Příklad 4.2
Výzkumná agentura zjišťovala u 128 náhodně vybraných osob, zda v referendu souhlasili se vstupem ČR do Evropské unie a zda nyní, řadu let po vstupu, mají stejný názor. Získané údaje obsahuje následující tabulka. Lze považovat zjištěnou změnu názorů za významnou?

Názor v referendu	Nynější názor na členství v EU: kladný	záporný
kladný	48	26
záporný	14	40

Řešení:
Vzhledem k tomu, že v tomto případě v časovém odstupu zjišťujeme hodnoty téže alternativní proměnné, dostáváme zvláštní typ čtyřpolní kontingenční tabulky. Na jejím základě provedeme výpočet hodnoty McNemarovy statistiky

$$Q_{MN} = \frac{(n_{12} - n_{21})^2}{n_{12} + n_{21}},$$

která má pro $n_{12} + n_{21} \geq 30$ asymptoticky chí-kvadrát rozdělení s jedním stupněm volnosti. Při 5% hladině významnosti tvoří kritický obor hodnoty testového kritéria převyšující 95% kvantil tohoto rozdělení (tj. 3,84). Z tabulky dostáváme hodnotu testového kritéria

$$Q_{MN} = \frac{(26-14)^2}{26+14} = \frac{144}{40} = 3{,}60,$$

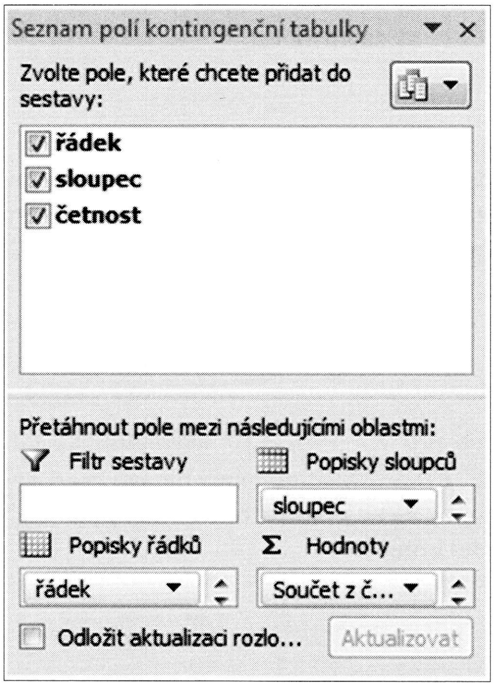
která nás neopravňuje na 5% hladině významnosti zamítnout testovanou hypotézu. Významnou změnu názorů jsme tedy neprokázali.

Excel

Při sestavení tabulky v Excelu nyní vyjdeme z četností, jež jsou součástí zadání příkladu. Kombinacím kategorií obou veličin přiřadíme příslušnou četnost.

	A	B	C	D
1	řádek	sloupec	četnost	
2	kladný	kladný	48	
3	kladný	záporný	26	
4	záporný	kladný	14	
5	záporný	záporný	40	
6				

Nástroje pro sestavení kontingenční tabulky používáme stejně jako v předchozím příkladu. Po jejich zobrazení jméno proměnné *řádek* přetáhneme do pole *Popisky řádků*, jméno proměnné *sloupec* do pole *Popisky sloupců*. Jméno proměnné *četnosti* přetáhneme do pole *Hodnoty*. Ještě je ovšem třeba provést změnu nastavení pole hodnot, která je zatím *Počet z četnosti*, na *Součet z četnosti* (klikneme na šipku, která je v pravé části tlačítka *Počet z četnosti* a dále zvolíme *Nastavení polí hodnot*). Obdržíme čtyřpolní kontingenční tabulku.



Součet z četnost	Sloupce		
Řádky	kladný	záporný	Celkem
kladný	48	26	74
záporný	14	40	54
Celkem	62	66	128

Výpočet testového kritéria provedeme z četností v tabulce standardním způsobem na základě výše uvedeného vzorce. (Pro výpočty je výhodnější zkopírovat si potřebné četnosti mimo tabulku, vzorce jsou přehlednější). Kritickou hodnotu určíme pomocí funkce CHISQ.INV.RT s argumenty 0,05 a 1; kvantil je, jak již bylo uvedeno, 3,84.

□

Příklad 4.3

Na základě průzkumu, provedeného u čtenářů časopisů A, B a C, byla sestavena následující kontingenční tabulka. Rozhodněte, zda výběr časopisu závisí na vzdělání čtenáře.

	Časopis			
Vzdělání	A	B	C	Celkem
ZŠ	75	75	50	200
SŠ	40	70	40	150
VŠ	35	5	10	50
Celkem	150	150	100	400

Řešení:

Excel

Sestavíme kontingenční tabulku.

Součet z četnost	**Časopi**			
Vzdělání	**A**	**B**	**C**	**Celkem**
základní	75	75	50	200
středoškolské	40	70	40	150
vysokoškolské	35	5	10	50
Celkem	**150**	**150**	**100**	**400**

Dále provedeme výpočet statistiky G.
Absolutní četnosti z kontingenční tabulky překopírujeme například do oblasti A1:D4.
Na jejich základě spočteme očekávané četnosti:
1) do pole A7 vložíme vzorec =A$4*$D$1/400;
2) do pole A8 vložíme vzorec =A$4*$D$2/400;
3) do pole A9 vložíme vzorec =A$4*$D$3/400;
pole zkopírujeme do sloupců B a C. Kontrolní součty všech sloupců i řádků a také celkový součet musí být stejné jako u původní kontingenční tabulky.
V oblasti A7:D10 získáme tedy následující tabulku

75	75	50	200
56,25	56,25	37,5	150
18,75	18,75	12,5	50
150	150	100	400

Pro účely výpočtu hodnoty statistiky G vložíme dále například do pole A11 vzorec =(A1-A7)^2/A7. Zkopírujeme jej do celé plochy kontingenční tabulky (A11:C13) a všechny obdržené hodnoty sečteme. Výsledkem je hodnota statistiky G, tedy 32,889.

0	0	0	
4,694444	3,361111	0,166667	
14,08333	10,08333	0,5	
			32,88889

Kritickou hodnotu určíme pomocí funkce CHISQ.INV.RT s argumenty 0,05 a 4, kvantil je 9,5. Na hladině významnosti 0,05 je závislost volby časopisu na vzdělání prokázána.
P-hodnota testu spočtená pomocí funkce CHISQ.TEST na základě výše uvedených zjištěných a očekávaných četností je prakticky nulová (1,3E-06), samozřejmě to opět znamená, že alternativní hypotéza byla prokázána.

◘

Cvičení

1. Na základě údajů v následující kontingenční tabulce (případně v souboru Zmena.xlsx) ověřte (na 5% hladině významnosti), zda ochota přestěhovat se do jiného města (1 = „naprostý souhlas“, 2 = „souhlas“, 3 = „je mi to jedno“, 4 = „nesouhlas“, 5 = „naprostý nesouhlas“) závisí na pohlaví (1 = muž, 2 = žena). Která pole v kontingenční tabulce nejvíce přispívají k hodnotě testového kritéria? Určete výběrový koeficient kontingence.

	Ochota přestěhovat se				
Pohlaví	1	2	3	4	5
1	4	35	15	28	48
2	14	23	19	39	50

2. Následující kontingenční tabulka je výsledkem třídění výběrových údajů o úrovni znalostí o Evropské unii (1 = „nevím nic“, 2 = „vím dost málo“, 3 = „vím poměrně dost“, 4 = „vím toho mnoho“ a stáří respondenta (pět věkových skupin, 1 = nejmladší, …, 5 = nejstarší). Lze říci, že úroveň znalostí o Evropské unii závisí na věku? Jsou splněny podmínky užití testu chí-kvadrát?

Úroveň znalostí	Věková skupina				
	1	2	3	4	5
1	2	2	4	7	6
2	13	20	13	31	40
3	27	36	50	40	50
4	10	10	12	13	14

3. Pro řídké tabulky s malými četnostmi se často doporučuje spojování kategorií proměnných. Spojte v předchozí tabulce první tři věkové kategorie do jedné („mladší osoby“) a zbývající věkové skupiny do druhé („starší osoby“) a přesvědčte se o tom, že takový postup může ovlivnit výsledek testu.

4. Při zjišťování spokojenosti studentů ekonomie se studiem jsme zjistili, že u 46 dotázaných studium splňuje očekávání, 22 od studia očekávalo více, 34 dotázaných mělo nižší nároky. Lze na 5% hladině významnosti říci, že je významný rozdíl mezi očekáváními a úrovní studia ekonomie?

Výsledky

1. $G = 9{,}566$; kritická hodnota pro α = 0,05 je 9,488 (závisí); naprostý souhlas a souhlas u obou pohlaví; C = 0,183; V = 0,187.

2. $G = 14{,}300$; kritická hodnota pro α = 0,05 je 21,026 (nezávisí); 4 očekávané četnosti z 20 (tj. 20 %) jsou nižší než pět, podmínky testu jsou splněny.

3. $G = 9{,}552$; kritická hodnota pro α = 0,05 je 7,815 (závisí).

4. Q_{MC} = 2,571; kritická hodnota pro α = 0,05 je 3,841 (nelze).

KAPITOLA V

ANALÝZA ROZPTYLU

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$$

$$F = \frac{\dfrac{S_{y,m}}{k-1}}{\dfrac{S_{y,v}}{n-k}}$$

5 A

Příkl
U 18
nán ú
li dru
hladi

Řešen
V **Tal**
kredit
k ově
analý
(podl
měrn
noven
noty p

Defin
motn
rozpt

mezis

vnitro

5 Analýza rozptylu

Příklad 5.1

U 18 studentů byl zjišťován počet získaných kreditů za poslední semestr a zaznamenán údaj o studované fakultě, viz **Tab. 5.1**. Rozhodněte pomocí vhodného testu, zda-li druh studované fakulty ovlivňuje počet získaných kreditů. Test proveďte na 5% hladině významnosti.

Tab. 5.1

Fakulta	Získané kredity
1	27,27,27,28,30,31
2	21,20,19,20,18,21
3	36,38,34,35,33,32

Řešení:

V **Tab. 5.1** jsou uspořádány hodnoty kvantitativní proměnné Y (počet získaných kreditů) podle hodnot faktoru X (studovaná fakulta). Vzhledem k povaze dat bude k ověření, zda-li je počet získaných kreditů ovlivněn studovanou fakultou, využita analýza rozptylu. Proměnná Y nabývá n hodnot, které je možné roztřídit do k skupin (podle variant faktoru X), n_i představuje počet pozorování v i-té skupině, $\overline{y}_i$ je průměrná hodnota proměnné Y v i-té skupině a $\overline{y}$ je celkový průměr proměnné Y, stanovený na základě všech n hodnot. Dále označme jednotlivé skupinové střední hodnoty proměnné Y symboly μ_i.

Definujme jednotlivé součty čtverců, které budeme využívat jednak k výpočtu samotného testového kritéria analýzy rozptylu, jednak k ověření předpokladů analýzy rozptylu. Celkový součet čtverců jako

$$S_y = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (y_{ij} - \overline{y})^2 ,$$

meziskupinový součet čtverců

$$S_{y,m} = \sum_{i=1}^{k} (\overline{y}_i - \overline{y})^2 n_i ,$$

vnitroskupinový součet čtverců

$$S_{y,v} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 .$$

Pro výše uvedené vztahy platí, že celkový součet čtverců je možné vypočítat jako součet meziskupinového a vnitroskupinového součtu čtverců, tj.

$$S_y = S_{y,m} + S_{y,v} .$$

Při analýze rozptylu testujeme hypotézu o rovnosti středních hodnot, proti alternativní hypotéze, která představuje negaci testované hypotézy, tedy

H_0: $\mu_1 = \mu_2 = \ldots = \mu_k$,
H_1: non H_0.

Testovým kritériem je statistika F, která se stanoví podle vzorce

$$F = \frac{\dfrac{S_{y,m}}{k-1}}{\dfrac{S_{y,v}}{n-k}},$$

kde $S_{y,m}$ a $S_{y,v}$ jsou výše definované součty čtverců.

Při platnosti testované hypotézy H_0 má toto testové kritérium F-rozdělení s $(k - 1)$ a $(n - k)$ stupni volnosti. Kritický obor je definován na základě nerovnosti

$$W_\alpha = \{F \geq F_{1-\alpha}(k - 1, n - k)\},$$

kde $F_{1-\alpha}$ představuje kvantil F-rozdělení. Pokud je hodnota testového kritéria větší než uvedený kvantil F-rozdělení, je zřejmé, že testovaná hypotéza o rovnosti středních hodnot bude na zvolené hladině významnosti α zamítnuta.

Pokud se na dané hladině významnosti podaří prokázat, že jednotlivé střední hodnoty proměnné Y nejsou shodné, tj. pokud je na dané hladině významnosti zamítnuta testovaná hypotéza o jejich rovnosti, dále se zkoumá těsnost této závislosti. Měříme ji pomocí poměru determinace, který je definován jako

$$P^2 = \frac{S_{y,m}}{S_Y} .$$

Tento poměr nabývá hodnot z uzavřeného intervalu od 0 do 1. Pokud je výsledkem hodnota 0, je to způsobeno tím, že meziskupinový součet čtverců je nulový. V takovém případě je celková variabilita tvořena pouze variabilitou uvnitř skupin a tedy proměnné jsou nezávislé (střední hodnoty proměnné Y jsou ve všech k skupinách podle faktoru X stejné). Čím je hodnota poměru determinace bližší jedné, tím je těsnost závislosti proměnné Y na faktoru X silnější.
Uvedené použití analýzy rozptylu vychází z předpokladu, že hodnoty proměnné Y v každé z k skupin představují výběry z normálního rozdělení, a že tyto výběry jsou nezávislé. Za závislé je možné považovat například výběry, kdy se sledují opakovaně hodnoty u stejných respondentů. Ověření normality lze provádět např. pomocí některé z grafických metod. Jak se často uvádí, porušení předpokladu normality proměnné Y nemá zásadní vliv na rozdělení statistiky F u analýzy rozptylu. Dalším předpokladem analýzy rozptylu je shoda všech skupinových rozptylů σ_i^2. Tento předpoklad je možné ověřit pomocí tzv. Bartlettova testu. V případě, že jsou rozsahy skupin stejné, uvádí se, že nesplnění předpokladu o shodě skupinových rozptylů nemá zásadní vliv na analýzu rozptylu.

V Bartlettově testu je ověřována shoda všech skupinových rozptylů proměnné Y proti alternativní hypotéze, která je negací testované hypotézy, tedy:

$$H_0\colon \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2,$$
$$H_1\colon \text{non } H_0.$$

Testovým kritériem je statistika

$$T = \frac{(n-k)\cdot \ln S^2 - \sum_{i=1}^{k}(n_i - 1)\cdot \ln s_i^2}{1 + \frac{1}{3\cdot(k-1)}\cdot\left(\sum_{i=1}^{k}\frac{1}{n_i - 1} - \frac{1}{n-k}\right)},$$

kde

$$s_i^2 = \frac{\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2}{n-1}$$

a

$$S^2 = \frac{S_{y,y}}{n-k}.$$

Toto testové kritérium má při platnosti testované hypotézy H_0 rozdělení χ^2 s $(k-1)$ stupni volnosti. Kritický obor je definován na základě nerovnosti

$$W_\alpha = \{\chi^2 \geq W_\alpha = \left\{\chi^2 \geq \chi^2_{1-\alpha}(k-1)\right\},$$

kde $\chi^2_{1-\alpha}$ představuje kvantil rozdělení chí-kvadrát. Pokud je hodnota testového kritéria větší než uvedený kvantil chí-kvadrát rozdělení, je zřejmé, že testovaná hypotéza o rovnosti skupinových rozptylů bude na hladině významnosti α zamítnuta a předpoklad analýzy rozptylu o shodně skupinový rozptylů nebude splněn.

V našem příkladu nejprve ověřme, zda-li je splněn předpoklad užití analýzy rozptylu pomocí Bartlettova testu. Jak bylo uvedeno výše, v našem konkrétním příkladu by však toto ověření nebylo bezpodmínečně nutné, neboť počty hodnot ve všech skupinách jsou stejné. Pomocné výpočty jsou uvedeny v Tab. 5.2.

Tab. 5.2

j	y_{ij}	i	n_i	$\overline{y}_i$	$(y_{ij}-\overline{y}_i)^2$	$(\overline{y}_i-\overline{y})^2 \cdot n_i$
1	27	1	6	28,3333	1,7778	3,1296
2	27				1,7778	
3	27				1,7778	
4	28				0,1111	
5	30				2,7778	
6	31				7,1111	
7	21	2	6	19,8333	1,3611	362,9630
8	20				0,0278	
9	19				0,6944	
10	20				0,0278	
11	18				3,3611	
12	21				1,3611	
13	36	3	6	34,6667	1,7778	298,6852
14	38				11,1111	
15	34				0,4444	
16	35				0,1111	
17	33				2,7778	
18	32				7,1111	
Σ	497	-	18	27,6111	45,5000	664,7778

Pro stanovení testového kritéria Bartlettova testu je třeba určit další hodnoty, které jsou uspořádány do Tab. 5.3.

Tab. 5.3

i	s_i^2	n_i	$\ln s_i^2$	$(n_i - 1)\cdot \ln s_i^2$	$\dfrac{1}{n_i - 1}$
1	3,0667	6	1,1206	5,6030	0,2
2	1,3667	6	0,3124	1,5619	0,2
3	4,6667	6	1,5404	7,7022	0,2
Σ	-	18	-	14,8671	0,6

S využitím Tab. 5.2 stanovme hodnotu

$$S^2 = \frac{45,50}{18-3} = 3,0333,$$

kterou budeme potřebovat pro výpočet testového kritéria. Výpočet samotného testového kritéria s využitím pomocných výpočtů z Tab. 5.2 a Tab. 5.3 je následující

$$T = \frac{(18-3)\cdot \ln(3,0333) - 14,8671}{1 + \dfrac{1}{3\cdot(3-1)}\cdot\left(0,6 - \dfrac{1}{18-3}\right)} = 1,6327.$$

Kritickou hodnotou pro Barttlettův test je v našem případě kvantil $\chi^2_{0,95}(2) = 5,9915$. Vzhledem k tomu, že hodnota testového kritéria nespadá do kritického oboru, na 5% hladině významnosti nezamítáme testovanou hypotézu Bartlettova testu, tedy předpoklad o rovnosti skupinových rozptylů můžeme považovat za splněný.

Přistupme nyní k testování hypotézy o rovnosti středních hodnot u tří fakult, tedy

H_0: $\mu_1 = \mu_2 = \mu_3$,
H_1: non H_0.

Testové kritérium F se určí s využitím pomocných výpočtů v Tab. 5.2 jako

$$F = \frac{\dfrac{664,7778}{3-1}}{\dfrac{45,5}{18-3}} = 109,5788.$$

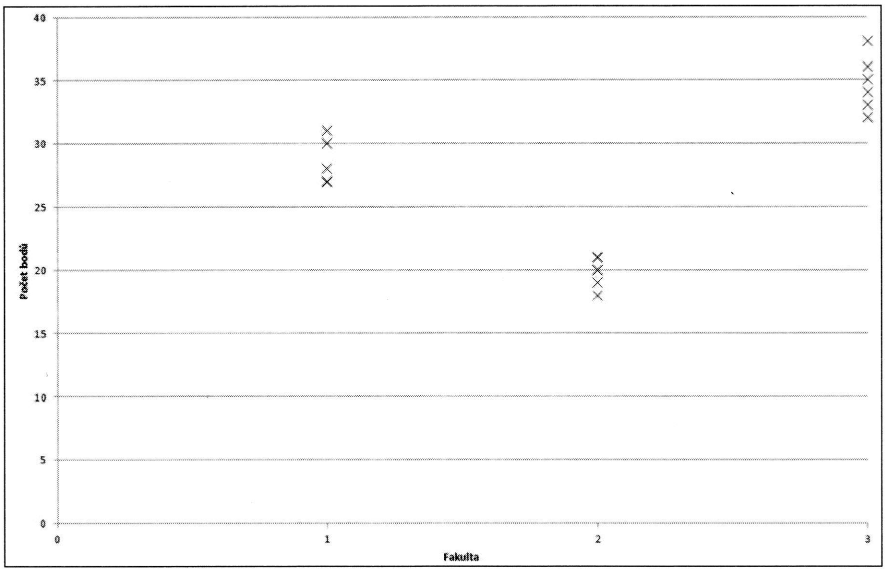
Kritickou hodnotou je kvantil $F_{0,95}(2,15) = 3,6823$. Protože hodnota testového kritéria je větší než kritická hodnota, na 5% hladině významnosti zamítáme testovanou hypotézu o rovnosti středních hodnot a prokázali jsme, že existuje alespoň jedna dvojice středních hodnot počtu získaných kreditů, která je odlišná. Protože jsme prokázali, že počet získaných kreditů je ovlivněn typem studované fakulty, stanovme těsnost závislosti pomocí poměru determinace

$$P^2 = \frac{664,7778}{710,2778} = 0,9359.$$

Z vypočteného poměru determinace je zřejmé, že se jedná o značně těsnou závislost.

Excel

Z grafu na Obr. 5.1 jsou patrné počty bodů jednotlivých studentů v rámci fakult. Je zřejmé, že počty bodů, dosahované u studentů ze druhé fakulty, jsou výrazně nižší než počty bodů studentů ze třetí fakulty.



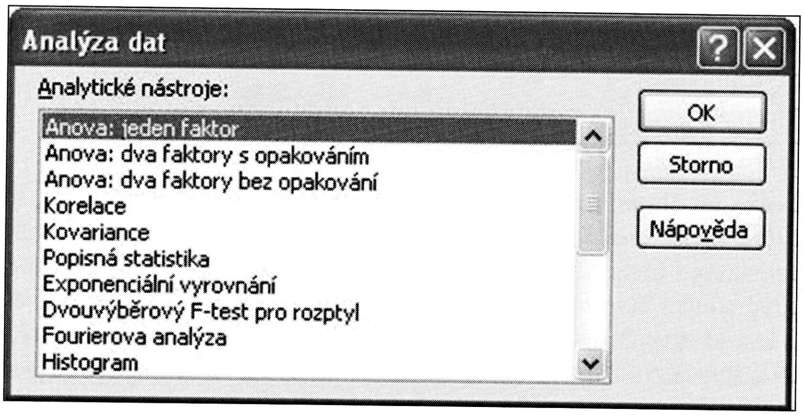
Obr. 5.1

Ukažme si nyní, jak vyřešíme příklad za pomoci Excelu. Vyvolání nástroje jednofaktorové analýzy rozptylu je následující:

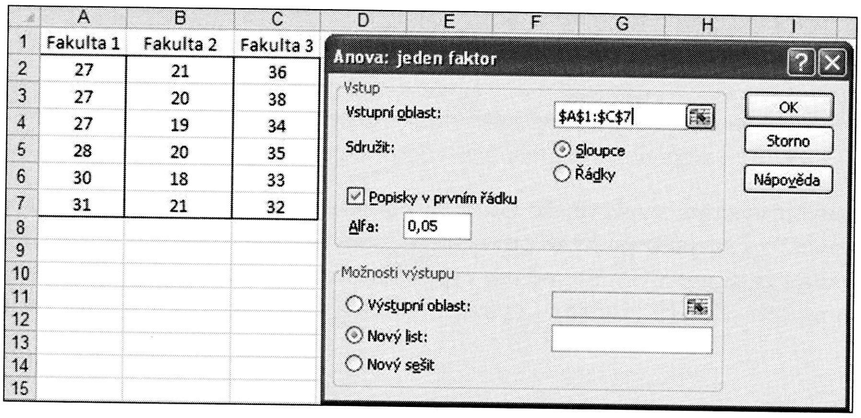
Data

Analýza dat

Anova: jeden faktor



Hodnoty kvantitativní proměnné *Y* máme uspořádány do jednotlivých sloupců, kterých je celkem *k*, tj. podle počtu variant faktoru *X*. V případě, že vybereme oblast, která obsahuje zároveň názvy jednotlivých variant faktoru *X*, tj. v našem případě označení jednotlivých fakult, musíme tuto skutečnost vyznačit ve vstupním okně, tj. zaškrtneme *Popisky v prvním řádku*. Do políčka *Alfa* uvedeme zvolenou hladinu významnosti, standardně je uvedena hodnota *0,05*. Pokud nezvolíme jinak, standardně je výstup ANOVA umístěn na nový list. V případě, že bychom chtěli konkrétní umístění v rámci aktivního listu, zvolíme *Výstupní oblast*, případně můžeme výstup umístit do *Nového sešitu*.



Výstup analýzy rozptylu je rozdělen do dvou tabulek. První tabulka obsahuje popisné charakteristiky kvantitativní proměnné Y (počet pozorování, součet hodnot, průměrnou hodnotu a rozptyl) pro všech k skupin.

Faktor				
Výběr	*Počet*	*Součet*	*Průměr*	*Rozptyl*
Fakulta 1	6	170	28,33333	3,066667
Fakulta 2	6	119	19,83333	1,366667
Fakulta 3	6	208	34,66667	4,666667

Druhá tabulka obsahuje standardizovaný výstup, popisující průběh testu, který jsme popsali výše. Ve sloupci *SS* jsou uvedeny jednotlivé součty čtverců. Označení *Mezi výběry* představuje meziskupinový součet čtverců, *Všechny výběry* představuje vnitroskupinový součet čtverců. Sloupec označený jako *Rozdíl* představuje stupně volnosti, tj. $(k - 1)$ a $(n - k)$. Ve sloupci *MS* jsou uvedeny podíly příslušného součtu čtverců *SS* a stupňů volnosti *Rozdíl.* Hodnota ve sloupci *F* představuje hodnotu testového kritéria definovanou výše, tj. jedná se o podíl dvou hodnot ze sloupce *MS.* V posledním sloupci této tabulky je uvedena kritická hodnota (pro zvolenou hladinu významnosti) daného testu, tj. kvantil F-rozdělení s příslušnými stupni volnosti. U počítačových výstupů bývá uvedena ještě tzv. p-hodnota, na jejímž základě můžeme přijmout závěr daného testu, aniž bychom museli porovnávat hodnotu testového kritéria s kritickou hodnotou. Vždy platí, že pokud je p-hodnota menší než zvolená hladina významnosti, testovanou hypotézu na dané hladině významnosti zamítáme. V našem případě je zřejmé, že testovanou hladinu o rovnosti středních hodnot na 5% hladině zamítáme, neboť p-hodnota je velmi malé číslo.

ANOVA						
Zdroj variability	*SS*	*Rozdíl*	*MS*	*F*	*Hodnota P*	*F krit*
Mezi výběry	664,7778	2	332,3889	109,5788	1,1204E-09	3,68232
Všechny výběry	45,5	15	3,033333			
Celkem	710,2778	17				

Z uvedených výstupů vyplývá, že hodnotu poměru determinace bychom museli stanovit „ručně“ a to jako podíl dvou výše popsaných součtů četverců, tj. jako podíl dvou hodnot ze sloupce *SS.* Stejně tak výpočet testového kritéria T u Bartlettova testu bychom museli stanovit „ručně“ s využitím dílčích výpočtů z uvedených výstupů.

Cvičení

1. Ve 4 lokalitách bylo náhodně osloveno celkem 28 respondentů, u nichž bylo zjišťováno, kolikrát navštíví hypermarket potravin v rámci jednoho měsíce. Pomocí vhodného testu rozhodněte, zda je počet návštěv hypermarketu ovlivněn lokalitou hypermarketu. Test proveďte na 5% hladině významnosti.

Lokalita	Počet návštěv
A	2,3,3,2,4,3,3
B	4,5,4,4,5,4,5
C	2,3,3,2,1,2,2
D	3,3,2,3,3,4,3

2. Ve 3 různých skupinách byl proveden test znalostí o Evropské unii. Každá ze skupin obsahovala 8 osob, u nichž byl zaznamenán počet bodů, získaný z daného testu. Výstup jednofaktorové analýzy rozptylu z MS Excel je uveden v následující tabulce. Na 5% hladině významnosti ověřte, zda se střední hodnoty počtu bodů z testu významně odlišují. Pokud ano, stanovte koeficient, který měří těsnost této závislosti, a interpretujte jej.

ANOVA						
Zdroj variability	*SS*	*Rozdíl*	*MS*	*F*	*Hodnota P*	*F krit*
Mezi výběry	785,3333	2	392,6667	28,19145	1,12857E-06	3,4668
Všechny výběry	292,5	21	13,92857			
Celkem	1077,833	23				

3. Pomocí experimentu byla ověřována spotřeba automobilů 5 různých typů. V každé ze skupin byla spotřeba měřena u 5 automobilů. Uvažujte tabulku, která obsahuje výstup jednofaktorové analýzy rozptylu z MS Excel. Doplňte chybějící údaje do tabulky a pomocí vhodného testu na 5% hladině významnosti rozhodněte, zda jsou střední hodnoty spotřeby automobilů významně odlišné. Pokud ano, stanovte koeficient těsnosti závislosti.

ANOVA					
Zdroj variability	*SS*	*Rozdíl*	*MS*	*F*	*F krit*
Mezi výběry					2,866081402
Všechny výběry	0,644	20	0,0322		
Celkem	14,4				

Výsledky

1. $F = 16{,}33$; p-hodnota = 5,36E-06; testovanou hypotézu o rovnosti středních hodnot zamítáme na 5% hladině významnosti (i na 1% hladině významnosti), a tak jsme prokázali, že se střední hodnoty počtu návštěv hypermarketu významně liší. Hodnota poměru determinace je $P^2 = 0{,}6712$, tj. jedná se o středně silnou intenzitu závislosti. Testovaná hypotéza Bartlettova testu je zamítnuta (na 1% i 5% hladině významnosti) a tedy předpoklad o shodě skupinových rozptylů je splněn.

2. $F = 28{,}19$; protože $F >$ „F $krit$", můžeme na 5% hladině významnosti testovanou hypotézu o rovnosti středních hodnot zamítnout, a tím pádem jsme prokázali, že se střední hodnoty počtu bodů z testu znalostí o EU významně liší. Hodnota poměru determinace je $P^2 = 0{,}7286$, což znamená, že se jedná o poměrně silnou intenzitu závislosti.

3. Po dopočtení bychom získali následující tabulku, kde je pro úplnost uvedena i p-hodnota, kterou bychom jednoduchým ručním výpočtem nestanovili.

ANOVA						
Zdroj variability	*SS*	*Rozdíl*	*MS*	*F*	*Hodnota P*	*F krit*
Mezi výběry	13,756	4	3,439	106,8012	3,37754E-13	2,866081
Všechny výběry	0,644	20	0,0322			
Celkem	14,4	24				

Kritická hodnota F $krit$ představuje kvantil $F_{0,95}(4,20) = 2{,}866$. Protože hodnota testového kritéria $F > F_{0,95}$, testovanou hypotézu o rovnosti středních hodnot spotřeby jednotlivých typů automobilů zamítáme a prokázali jsme, že se tyto střední hodnoty na 5% hladině významnosti odlišují. Hodnota poměru determinace je $P^2 = 0{,}9553$, a jedná se tedy o velmi silnou intenzitu závislosti.

KAPITOLA VI

REGRESNÍ A KORELAČNÍ ANALÝZA

$$b_1 = \frac{n\sum_{i=1}^{n} y_i x_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2} = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\overline{x^2} - \overline{x}^2}$$

6

6.1

Přík
U 13
vy n
vyjá
níhc
střed
uspo

Dob	
Poč	

Řeše
Je m
čet d
vého
vhoc

kde
(nez

Regr
získa

Pro c
Para
stant

6 Regresní a korelační analýza

6.1 Jednoduchá regrese

Příklad 6.1

U 13 náhodně vybraných studentů byla pomocí experimentu zjišťována doba přípravy na určitý test (v minutách) a počet dosažených bodů. Pomocí regresní přímky vyjádřete závislost počtu bodů na době přípravy studenta. Zhodnoťte kvalitu regresního modelu, vyjádřete intenzitu závislosti počtu bodů na době přípravy. Odhadněte střední hodnotu počtu bodů studenta, který se připravoval 182 minut. Údaje jsou uspořádány do Tab. 6.1.

Tab. 6.1

Doba př.	160	160	162	163	161	170	172	177	179	178	182	184	183
Počet b.	57	55	59	60	52	67	69	74	75	76	78	80	87

Řešení:

Je možné očekávat, že s rostoucí dobou přípravy studenta bude v průměru růst i počet dosažených bodů z testu. Tuto skutečnost je možné vyčíst i z následujícího bodového diagramu, viz Obr. 6.1. Na základě tohoto grafu je také možné usuzovat, že vhodným tvarem bude lineární model. Jeho tvar můžeme vyjádřit pomocí vztahu

$$y = \beta_0 + \beta_1 x + \varepsilon,$$

kde y jsou hodnoty vysvětlované (závislé) proměnné, x jsou hodnoty vysvětlující (nezávislé) proměnné a ε je nesystematická (náhodná) složka.

Regresní přímka $\eta = \beta_0 + \beta_1 x$ vyjadřuje lineární vztah mezi střední hodnotou počtu získaných bodů a dobou přípravy.

Pro odhad parametrů regresní přímky β_0 a β_1 využijeme metodu nejmenších čtverců. Parametr b_1 představuje směrnici regresní přímky a parametr b_0 představuje její konstantu. Vypočteme je pomocí následujících vzorců:

$$b_1 = \frac{n\sum_{i=1}^{n} y_i x_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2} = \frac{\overline{xy} - \overline{x}\cdot\overline{y}}{\overline{x^2} - \overline{x}^2},$$

resp.

$$b_0 = \overline{y} - b_1\overline{x}\,.$$

Ke zhodnocení kvality použitého regresního modelu se používá koeficient determinace R^2. K tomu, abychom jej mohli určit, je nejprve nutné definovat jednotlivé součty čtverců.

Teoretický součet čtverců definujme podle vzorce

$$S_{y,T} = \sum_{i=1}^{n} (Y_i - \overline{y})^2,$$

reziduální součet čtverců jako

$$S_{y,R} = \sum_{i=1}^{n} (y_i - Y_i)^2$$

a celkový součet čtverců podle vzorce

$$S_y = \sum_{i=1}^{n} (y_i - \overline{y})^2,$$

kde y_i jsou skutečně naměřené hodnoty vysvětlované proměnné Y, Y_i jsou očekávané (teoretické) hodnoty vysvětlované proměnné Y získané na základě modelu, které získáme dosazením hodnot x_i do odhadnutého modelu a $\overline{y}$ je aritmetický průměr skutečně naměřených hodnot vysvětlované proměnné Y.
Pro výše uvedené součty čtverců platí vztah

$$S_y = S_{y,R} + S_{y,T}\,.$$

Koeficient determinace je pak definován jako podíl

$$R^2 = \frac{S_{y,T}}{S_y} = 1 - \frac{S_{y,R}}{S_y},$$

nabývá hodnot z intervalu od 0 do 1 a po vynásobení stem je interpretován jako podíl variability hodnot vysvětlované proměnné v %, kterou se podařilo vysvětlit pomocí daného regresního modelu.

Poznámka: Pokud bychom porovnávali kvalitu regresních modelů s různým počtem parametrů (například regresní přímku s regresní parabolou), je třeba použít hodnotu upraveného koeficientu determinace, který se stanoví podle vzorce

$$R^2_{adj.} = 1 - (1 - R^2)\frac{n-1}{n-p}.$$

Upravený koeficient determinace zohledňuje počet regresních parametrů p daného regresního modelu a jeho interpretace je identická.

Ke zhodnocení modelu jako celku slouží tzv. **celkový F-test**. Testovaná hypotéza v tomto testu obsahuje tvrzení, že všechny regresní parametry β_j ($j = 1, \ldots, k$), kromě konstanty, jsou rovny nule, což znamená, že v modelu není ani jedna vysvětlující proměnná X_j, která je statisticky významná. Alternativní hypotéza popírá platnost tohoto tvrzení, tedy

H_0: $\beta_0 = c$; $\beta_1 = \beta_2 = \ldots = \beta_k = 0$,
H_1: non H_0.

Testovým kritériem je statistika F, která se stanoví podle vzorce

$$F = \frac{\dfrac{S_{y,T}}{p-1}}{\dfrac{S_{y,R}}{n-p}},$$

kde $p = k + 1$ je počet regresních parametrů a k je počet vysvětlujících proměnných.

Kritický obor je dán nerovností $W_\alpha = \{F; F \geq F_{1-\alpha}\}$, kde $F_{1-\alpha}$ představuje kvantil F-rozdělení s $(p - 1)$ a $(n - p)$ stupni volnosti.

Dílčí t-testy
K postupnému ověření významnosti konstanty a vysvětlující proměnné v modelu samostatně slouží dílčí t-testy.

Test hypotézy o parametru β_0:
Pomocí prvního dílčího t-testu otestujeme hypotézu o nulové hodnotě konstanty

H_0: $\beta_0 = 0$,
H_1: $\beta_0 \neq 0$.

Testovým kritériem je statistika t, která se vypočítá podle vzorce

$$t = \frac{b_0}{s(b_0)},$$

kde směrodatná chyba odhadu parametru $s(b_0)$ se stanoví podle vzorce

$$s(b_0) = s_R \cdot \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n \cdot \sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}},$$

a kde odmocnina z reziduálního rozptylu s_R se určí jako

$$s_R = \sqrt{\frac{S_{y,R}}{n-2}}.$$

Kritický obor je dán nerovností $W_\alpha = \{t; |t| \geq t_{1-\alpha/2}\}$, kde $t_{1-\alpha/2}$ představuje kvantil t-rozdělení s $(n-2)$ stupni volnosti.

Test hypotézy o parametru β_1:
Pomocí druhého dílčího t-testu ověřujeme existenci vztahu mezi vysvětlovanou proměnnou Y a vysvětlující proměnnou X.

H_0: $\beta_1 = 0$,
H_1: $\beta_1 \neq 0$.

Testovým kritériem je statistika t, která se vypočítá podle vzorce

$$t = \frac{b_1}{s(b_1)},$$

kde

$$s(b_1) = s_R \cdot \sqrt{\frac{n}{n \cdot \sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}}.$$

Kritický obor je opět dán nerovností $W_\alpha = \{t; |t| \geq t_{1-\alpha/2}\}$, kde $t_{1-\alpha/2}$ představuje kvantil t-rozdělení s $(n-2)$ stupni volnosti.

Intervaly spolehlivosti pro regresní parametry je možné určit podle vzorce

$$P(b_j - t_{1-\alpha/2}(n-2) \cdot s(b_j) < \beta_j < b_j + t_{1-\alpha/2}(n-2) \cdot s(b_j)) = 1-\alpha .$$

Intervaly spolehlivosti pro podmíněnou střední hodnotu vysvětlované proměnné Y_0 je možné určit podle vzorce

$$P(Y_0 - t_{1-\alpha/2}(n-2) \cdot s(Y_0) < \eta_0 < Y_0 + t_{1-\alpha/2}(n-2) \cdot s(Y_0)) = 1-\alpha ,$$

kde

$$s(Y_0) = s_R \cdot \sqrt{\frac{1}{n} + (x_0 - \overline{x})^2 \cdot \left(\sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n} \right)^{-1}} .$$

Intervaly spolehlivosti pro individuální (konkrétní) hodnoty y_0 je možné určit podle vzorce

$$P(Y_0 - t_{1-\alpha/2}(n-2) \cdot s(y_0) < Ey_0 < Y_0 + t_{1-\alpha/2}(n-2) \cdot s(y_0)) = 1-\alpha ,$$

kde

$$s(y_0) = s_R \cdot \sqrt{1 + \frac{1}{n} + (x_0 - \overline{x})^2 \cdot \left(\sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n} \right)^{-1}} .$$

K určení intenzity lineární závislosti slouží korelační koeficient, který lze stanovit podle vzorce

$$r_{yx} = r_{xy} = \sqrt{R^2} .$$

Tento koeficient nabývá hodnot z intervalu od – 1 do 1, přičemž znaménko korelačního koeficientu udává směr závislosti a odpovídá znaménku parametru b_1 u regresní přímky.

Ze zadání našeho příkladu vyplývá, že počet bodů z testu je vysvětlovaná proměnná Y a doba přípravy studenta je vysvětlující proměnná X. Pomocné výpočty uspořádejme do Tab. 6.2.

Poznámka: Při výpočtech jednotlivých regresních parametrů, testových kritérií a intervalů byla vždy použita všechna desetinná místa (nebylo zaokrouhlováno), a proto se mohou výsledné hodnoty nepatrně lišit od hodnot, které by byly stanoveny „ručním" výpočtem se zaokrouhlováním.

Tab. 6.2

i	x_i	y_i	$x_i y_i$	x_i^2
1	160	57	9 120	25 600
2	160	55	8 800	25 600
3	162	59	9 558	26 244
4	163	60	9 780	26 569
5	161	52	8 372	25 921
6	170	67	11 390	28 900
7	172	69	11 868	29 584
8	177	74	13 098	31 329
9	179	75	13 425	32 041
10	178	76	13 528	31 684
11	182	78	14 196	33 124
12	184	80	14 720	33 856
13	183	87	15 921	33 489
Σ	2 231	889	153 776	383 941

Výpočet paramterů regresní přímky podle výše uvedených vzorců je následující:

$$b_1 = \frac{13 \cdot 153776 - 2231 \cdot 889}{13 \cdot 383941 - 2231^2} = 1{,}13387\,,$$

$$b_0 = \frac{889}{13} - 1{,}13387 \cdot \frac{2231}{13} = -126{,}204\,.$$

Výsledný tvar regresní přímky můžeme zapsat do tvaru

$$Y = -126{,}204 + 1{,}13387 \cdot x\,.$$

Ke zhodnocení kvality daného regresního modelu proveďme rozklad celkového součtu čtverců na jednotlivé složky. Údaje uspořádejme do Tab. 6.3.

Tab. 6.3

i	x_i	y_i	Y_i	$(y_i - Y_i)^2$	$(y_i - \bar{y})^2$
1	160	57	55,2152	3,1855	129,6095
2	160	55	55,2152	0,0463	179,1479
3	162	59	57,4829	2,3015	88,0710
4	163	60	58,6168	1,9132	70,3018
5	161	52	56,3491	18,9144	268,4556
6	170	67	66,5539	0,1990	1,9172
7	172	69	68,8216	0,0318	0,3787
8	177	74	74,4910	0,2411	31,5325
9	179	75	76,7587	3,0931	43,7633
10	178	76	75,6249	0,1407	57,9941
11	182	78	80,1603	4,6671	92,4556
12	184	80	82,4281	5,8956	134,9172
13	183	87	81,2942	32,5560	346,5325
Σ	2 231	889	-	73,1853	1 445,0769

Z předchozí tabulky dostáváme, že jednotlivé součty čtverců jsou následující

$$S_y = 1445{,}0769\,,$$

$$S_{y,R} = 73{,}1853\,,$$

teoretický součet čtverců získáme podle výše uvedeného vztahu jako rozdíl celkového a reziduálního součtu čtverců

$$S_{y,T} = 1445{,}0769 - 73{,}1853 = 1371{,}8994\,.$$

Koeficient determinace pak určíme podle výše uvedeného vztahu jako podíl

$$R^2 = \frac{1371{,}8994}{1445{,}0769} = 1 - \frac{73{,}1853}{1445{,}0769} = 0{,}9494\,.$$

Z hodnoty vypočteného koeficientu determinace vyplývá, že pomocí daného regresního modelu se podařilo vysvětlit 94,94 % variability hodnot proměnné počet získaných bodů z testu.

K ověření vhodnosti modelu jako celku využijeme celkový F-test, jehož testovaná a alternativní hypotézy jsou

H_0: $\beta_0 = c$, $\beta_1 = 0$,
H_1: non H_0.

Testové kritérium stanovíme podle vzorce

$$F = \frac{\frac{1371,8994}{2-1}}{\frac{73,1853}{13-2}} = 206,20,$$

kritickou hodnotou je kvantil $F_{0,95}(1;11) = 4,84$. Vzhledem k tomu, že hodnota testového kritéria spadá do kritického oboru, na 5% hladině významnosti zamítáme testovanou hypotézu celkového F-testu a prokázali jsme, že existuje alespoň jedna vysvětlující proměnná (u přímky právě jedna), která je statisticky významná.

K ověření významnosti konstanty a vysvětlující proměnné v regresním modelu samostatně použijeme dílčí t-testy.

V případě parametru β_0 postupujeme následovně

H_0: $\beta_0 = 0$,
H_1: $\beta_0 \neq 0$.

Testovým kritériem je statistika t, která se určí jako

$$t = \frac{-126,204}{13,57} = -9,30,$$

kde směrodatná chyba $s(b_0)$ se stanoví podle vzorce

$$s(b_0) = 2,5794 \cdot \sqrt{\frac{383941}{13 \cdot 383941 - (2231)^2}} = 13,57,$$

a kde odmocnina z reziduálního rozptylu S_R se určí jako

$$s_R = \sqrt{\frac{73{,}1853}{13-2}} = 2{,}5794 .$$

Kritická hodnota je dána kvantilem $t_{0{,}975}(11) = 2{,}20$. Vzhledem k tomu, že hodnota testového kritéria spadá do kritického oboru, na 5% hladině významnosti zamítáme testovanou hypotézu o nulové hodnotě konstanty.

V případě parametru β_1 postupujeme následovně

H_0: $\beta_1 = 0$,
H_1: $\beta_1 \neq 0$.

Testovým kritériem je statistika t, která se určí jako

$$t = \frac{1{,}13387}{0{,}0789} = 14{,}3597 ,$$

kde směrodatná chyba $s(b_1)$ se stanoví podle vzorce

$$s(b_1) = 2{,}5794 \cdot \sqrt{\frac{13}{13 \cdot 383941 - (2231)^2}} = 0{,}0789 .$$

Kritická hodnota je dána kvantilem $t_{0{,}975}(11) = 2{,}20$. Vzhledem k tomu, že hodnota testového kritéria spadá do kritického oboru, na 5% hladině významnosti zamítáme testovanou hypotézu a prokázali jsme, že střední hodnota počtu bodů z testu statisticky významně závisí na délce přípravy studenta.

Intervaly spolehlivosti pro regresní parametry je možné určit podle vzorců

$$P(1{,}13387 - 2{,}20 \cdot 0{,}0789 < \beta_1 < 1{,}13387 + 2{,}20 \cdot 0{,}0789) = 0{,}95 ,$$
$$P(0{,}9603 < \beta_1 < 1{,}3075) = 0{,}95 ,$$

$$P(-126{,}204 - 2{,}20 \cdot 13{,}57 < \beta_0 < -126{,}204 + 2{,}20 \cdot 13{,}57) = 0{,}95 ,$$
$$P(-156{,}058 < \beta_0 < -96{,}3504) = 0{,}95 .$$

Interval spolehlivosti pro střední hodnotu počtu bodů studentů, kteří se připravovali 182 minut, je možné určit podle vzorce

$$P(80{,}1594 - 2{,}20 \cdot 1{,}0882 < \eta_{182} < 80{,}1594 + 2{,}20 \cdot 1{,}0882) = 0{,}95\,,$$
$$P(77{,}7654 < \eta_{182} < 82{,}5534) = 0{,}95\,,$$

kde

$$s(Y_0) = 2{,}5794 \cdot \sqrt{\frac{1}{13} + \frac{(182 - 171{,}6154)^2}{383941 - \frac{2231^2}{13}}} = 1{,}0882\,.$$

Interval spolehlivosti pro individuální (konkrétní) hodnotu studenta, který se připravoval 182 minut je možné určit podle vzorce

$$P(80{,}1594 - 2{,}20 \cdot 2{,}7995 < Ey_{182} < 80{,}1594 + 2{,}20 \cdot 2{,}7995) = 0{,}95\,,$$
$$P(74{,}0005 < Ey_{182} < 86{,}3183) = 0{,}95\,,$$

kde

$$s(y_0) = 2{,}5794 \cdot \sqrt{1 + \frac{1}{13} + \frac{(182 - 171{,}6154)^2}{383941 - \frac{2231^2}{13}}} = 2{,}7995\,.$$

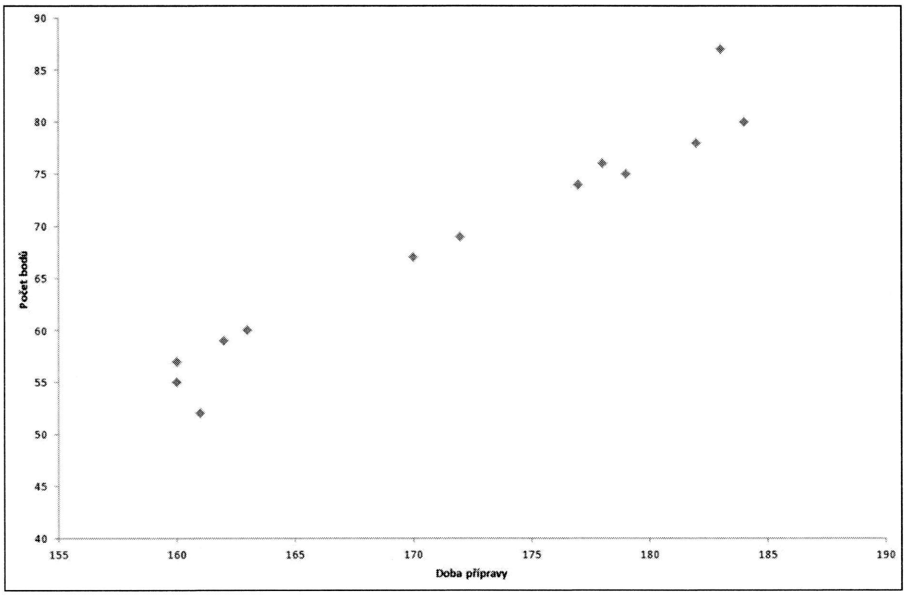
Hodnotu korelačního koeficientu, který měří intenzitu lineární závislosti počtu bodů na době přípravy studenta, vypočteme jako odmocninu z koeficientu determinace, přičemž znaménko je kladné (podle hodnoty parametru b_1), tedy

$$r_{yx} = \sqrt{R^2} = \sqrt{0{,}9494} = 0{,}9743\,.$$

Podle vypočtené hodnoty korelačního koeficientu usuzujeme, že počet bodů z testu je velmi silně přímo úměrně závislý na době přípravy.

Excel

V grafu na následujícím obrázku jsou znázorněny počty bodů v závislosti na době přípravy jednotlivých studentů. Je vidět, že s růstem doby přípravy jednotlivých studentů roste také počet získaných bodů. Jak bylo uvedeno výše, vhodným modelem k popsání této závislosti by mohla být regresní přímka.



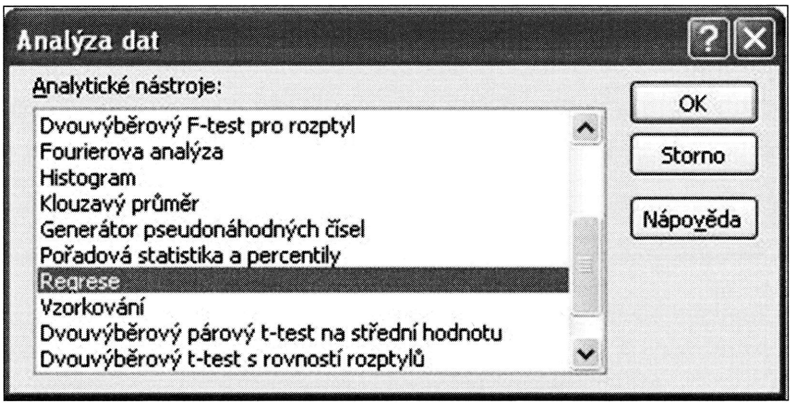
Obr. 6.1

Ukažme si nyní, jak vyvoláme regresní analýzu v Excelu.

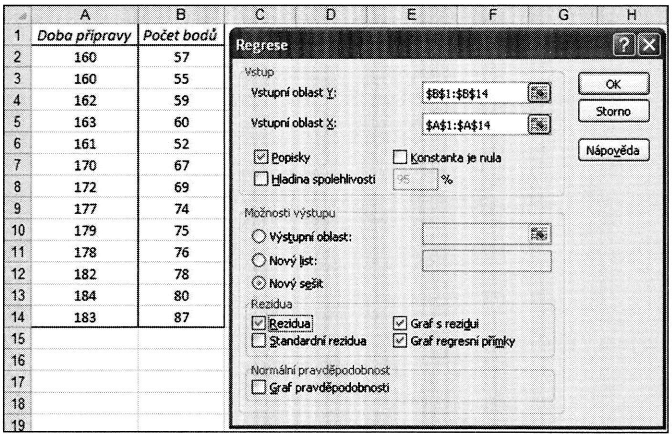
Data

Analýza dat

Regrese



Ve vstupním okně je třeba vyznačit oblast vysvětlované proměnné Y a vysvětlující proměnné X. Pokud označíme oblast včetně názvů proměnných, tuto skutečnost je nutné vyznačit zaškrtnutím ve volbě *Popisky*. Dále je třeba uvést *Hladinu spolehlivosti* pro příslušné intervaly spolehlivosti jednotlivých regresních parametrů. Standardně je uvedena hodnota *95 %*, tj. pokud tuto hodnotu nezměníme, budou ve výstupu uvedeny 95% intervaly spolehlivosti pro oba regresní parametry. Stejně jako v případě analýzy rozptylu je třeba zvolit umístění výstupu regresní analýzy. Standardně je nabízena volba *Nový list*, alternativou je umístění do zvolené *Výstupní oblasti* v rámci aktivního listu či umístění do *Nového sešitu*. Ve vstupním okně je ještě možné zaškrtnou různé grafické výstupy a hodnoty reziduí. Pokud zaškrtneme volbu *Rezidua*, do výstupu budou uvedeny jednak vyrovnané hodnoty, jednak jednotlivá rezidua. Zvolením položky *Graf regresní přímky* získáme graf napozorovaných a vyrovnaných hodnot.



Základní výstup regresní analýzy je rozdělen do tří tabulek. V první tabulce, označené jako *Regresní statistika*, jsou uvedeny následující hodnoty: korelační koeficient mezi vysvětlovanou a vysvětlující proměnnou r_{xy} pojmenovaný jako *Násobné R*, koeficient determinace R^2 označený názvem *Hodnota spolehlivosti R*, upravený koeficient determinace $R^2_{adj.}$ pojmenovaný jako *Nastavená hodnota spolehlivosti*, směrodatná chyba odhadu s_R, která je označena jako *Chyba stř. hodnoty* a počet pozorování.

Regresní statistika	
Násobné R	0,974348707
Hodnota spolehlivosti R	0,949355404
Nastavená hodnota spolehlivosti	0,94475135
Chyba stř. hodnoty	2,579382191
Pozorování	13

Druhá tabulka s názvem *ANOVA* obsahuje rozklad celkového součtu čtverců *Celkem* na část vysvětlenou regresním modelem, tj. teoretický součet čtverců označený jako *Regrese*, a zbytek, tj. reziduální součet čtverců označený jako *Rezidua*. Dále je v tabulce uveden test o modelu, tedy celkový F-test. Sloupec označený jako *Rozdíl* obsahuje příslušné stupně volnosti, tj. $(p-1)$ a $(n-p)$, kde p je počet regresních parametrů, tj. v případě regresní přímky dva. Sloupec *SS* obsahuje jednotlivé součty čtverců, *MS* obsahuje podíly jednotlivých součtů čtverců a příslušných stupňů volnosti. Hodnota ve sloupci *F* představuje testové kritérium celkového F-testu. *P*-hodnota tohoto testu je uvedena v posledním sloupci a je označena jako *Významnost F.*

ANOVA					
	Rozdíl	*SS*	*MS*	*F*	*Významnost F*
Regrese	1	1371,891586	1371,8916	206,1998753	1,80234E-08
Rezidua	11	73,18533737	6,6532125		
Celkem	12	1445,076923			

Z uvedené p-hodnoty vyplývá, že testovaná hypotéza celkového F-testu, je na 5%, ale i 1% hladině významnosti zamítnuta, neboť je tato hodnota menší než 0,05 resp. 0,01.

Třetí tabulka obsahuje jednak odhady regresních parametrů, jednak příslušné t-testy. Parametr b_0 je vždy označen jako *Hranice* a směrnice přímky, tj. parametr b_1 je označen názvem vysvětlující proměnné, pokud byl vložen, v našem případě *Doba přípravy*. Bodové odhady jednotlivých parametrů jsou uvedeny ve sloupci *Koeficienty*. Směrodatné chyby odhadu parametrů $s(b_j)$ jsou uvedeny ve sloupci *Chyba stř. hodnoty*. Ve sloupci, který je označen *t Stat*, jsou hodnoty testového kritéria dílčích *t*-testů. *P*-hodnota pro test o příslušném parametru je označena *Hodnota P*. Horní a dolní mez intervalu spolehlivosti pro regresní parametry pro zvolenou *Hladinu spolehlivosti*, uvedenou ve vstupním okně, v našem případě horní a dolní mez 95% intervalu spolehlivosti je možné najít v dalších sloupcích, tj. *Dolní 95%* a *Horní 95%.*

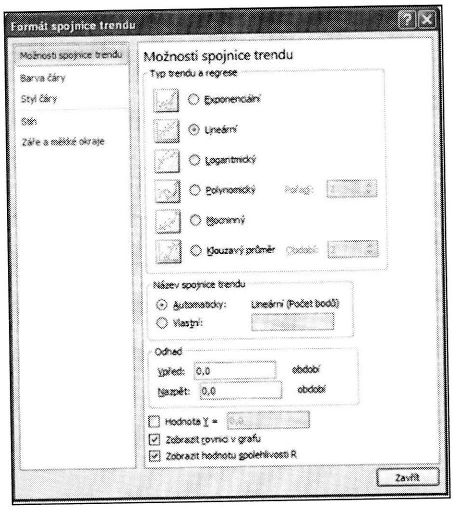
	Koeficienty	*Chyba stř. hodnoty*	*t Stat*	*Hodnota P*	*Dolní 95%*	*Horní 95%*	*Dolní 95,0%*	*Horní 95,0%*
Hranice	-126,2043685	13,56995485	-9,3002792	1,51869E-06	-156,0716378	-96,33709926	-156,0716378	-96,33709926
Doba přípravy	1,133866782	0,078961944	14,359661	1,80234E-08	0,960072715	1,307660849	0,960072715	1,307660849

Pokud bychom ve vstupním okně vybrali také rezidua, získali bychom následující tabulku, která obsahuje jednak *Očekávané počty bodů*, tj. vyrovnané hodnoty, a jednak hodnoty *Reziduí* odpovídající jednotlivým pozorováním.

Pro úpravu vytvořeného grafu (*Graf regresní přímky*) „klikněme“ pravým tlačítkem myši na libovolný napozorovaný (nikoliv vyrovnaný) počet bodů, zvolme možnost *Přidat spojnici trendu*, vyberme položku *Lineární*, dále vyberme *Zobrazit rovnici v grafu* a *Zobrazit hodnotu spolehlivosti R*, viz následující dialogové okno.

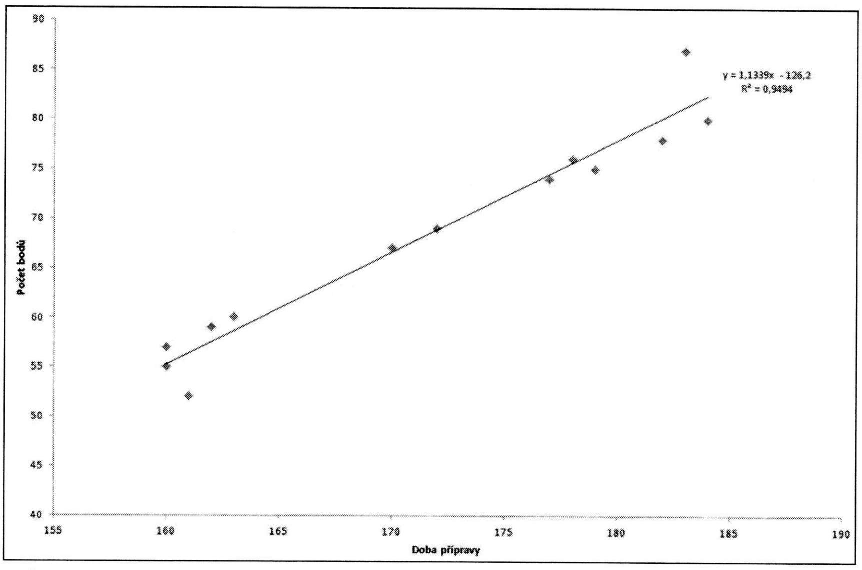
REZIDUA

Pozorování	Očekávané Počet bodů	Rezidua
1	55,21431661	1,785683391
2	55,21431661	-0,214316609
3	57,48205017	1,517949827
4	58,61591696	1,384083045
5	56,34818339	-4,348183391
6	66,55298443	0,447015571
7	68,82071799	0,179282007
8	74,4900519	-0,490051903
9	76,75778547	-1,757785467
10	75,62391869	0,376081315
11	80,15938581	-2,159385813
12	82,42711938	-2,427119377
13	81,2932526	5,706747405

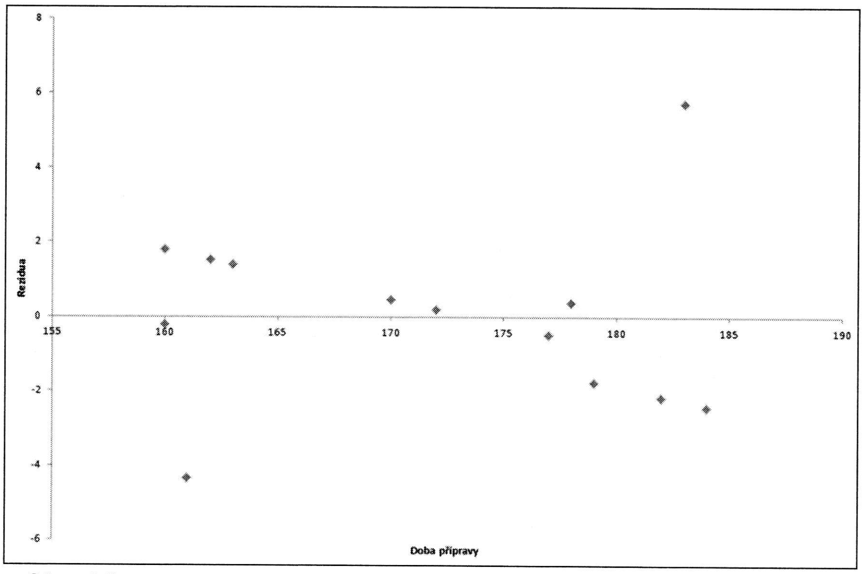


Tím získáme graf, ve kterém je kromě napozorovaných hodnot zobrazena také regresní přímka a její rovnice, včetně hodnoty koeficientu determinace, viz Obr. 6.2.

Poslední graf, který jsme si nechali vygenerovat, obsahuje hodnoty reziduí pro jednotlivá pozorování.



Obr. 6.2



Obr. 6.3

□

Příklad 6.2

U 13 pracovníků byla zjišťována chybovost případů při řešení obtížných úkolů (v procentech) v závislosti na délce jejich praxe (v měsících). Získané údaje byly uspořádány do Tab. 6.4. Pomocí regresní hyperboly vyjádřete závislost chybovosti pracovníka na délce jeho praxe. Zhodnoťte kvalitu regresního modelu a vyjádřete intenzitu této závislosti.

Tab. 6.4

Délka praxe	Chybovost případů
1	55,26
2	35,20
2	31,16
3	25,12
5	18,72
5	19,05
6	18,62
7	16,47
9	15,13
11	12,21
12	17,98
15	13,02
22	12,04

Řešení:

Při zkoumání vztahu mezi chybovostí při řešení obtížných úkolů je možné očekávat, že s růstem délky praxe bude docházet k jejímu poklesu. Tuto skutečnost je možné vyčíst i z následujícího bodového diagramu, viz Obr. 6.4.

Tvar regresní hyperboly je možné vyjádřit následujícím způsobem

$$\eta = \beta_0 + \beta_1 \frac{1}{x}.$$

Vzhledem k tomu, že zvolená regresní funkce je lineární z hlediska regresních parametrů, je možné její parametry, stejně jako v případě regresní přímky, odhadnout pomocí metody nejmenších čtverců. Při odhadování jejích parametrů se využívají výše uvedené vzorce pro regresní přímku, přičemž za hodnoty vysvětlující proměnné se dosazují převrácené hodnoty chybovosti pracovníků, tedy vzorce je možné upravit do tvaru

$$b_1 = \frac{n\sum_{i=1}^{n} y_i / x_i - (\sum_{i=1}^{n} 1/x_i)\sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} 1/x_i^2 - (\sum_{i=1}^{n} 1/x_i)^2},$$

resp.

$$b_0 = \frac{\sum_{i=1}^{n} y_i}{n} - b_1 \frac{\sum_{i=1}^{n} 1/x_i}{n}.$$

Pomocné výpočty uspořádejme do Tab. 6.5.

Tab. 6.5

i	$1/x_i$	y_i	y_i/x_i	$(1/x_i)^2$
1	1,00	55,26	55,2600	1,0000
2	0,50	35,20	17,6000	0,2500
3	0,50	31,16	15,5800	0,2500
4	0,33	25,12	8,3733	0,1111
5	0,20	18,72	3,7440	0,0400
6	0,20	19,05	3,8100	0,0400
7	0,17	18,62	3,1033	0,0278
8	0,14	16,47	2,3529	0,0204
9	0,11	15,13	1,6811	0,0123
10	0,09	12,21	1,1100	0,0083
11	0,08	17,98	1,4983	0,0069
12	0,07	13,02	0,8680	0,0044
13	0,05	12,04	0,5473	0,0021
Σ	3,44	289,98	115,5282	1,7734

Dosazením do výše uvedených vzorců získáme odhady parametrů regresní hyperboly

$$b_1 = \frac{13 \cdot 115{,}5282 - 3{,}44 \cdot 289{,}98}{13 \cdot 1{,}7734 - 3{,}44^2} = 44{,}9498,$$

$$b_0 = \frac{289{,}98}{13} - 44{,}9498 \cdot \frac{3{,}44}{13} = 10{,}4106.$$

Výsledný tvar regresní hyperboly můžeme zapsat jako

$$Y = 10{,}4106 + \frac{44{,}9498}{x}.$$

Ke zhodnocení kvality tohoto regresního modelu pomocí koeficientu determinace je třeba stanovit další pomocné výpočty, které jsou uspořádány do Tab. 6.6.

Tab. 6.6

i	$1/x_i$	y_i	Y_i	$(y_i - Y_i)^2$	$(y_i - \bar{y})^2$
1	1,00	55,26	55,3604	0,0101	1 085,9560
2	0,50	35,20	32,8855	5,3569	166,2513
3	0,50	31,16	32,8855	2,9774	78,3906
4	0,33	25,12	25,3939	0,0750	7,9177
5	0,20	18,72	19,4006	0,4632	12,8605
6	0,20	19,05	19,4006	0,1229	10,6025
7	0,17	18,62	17,9022	0,5152	13,5877
8	0,14	16,47	16,8320	0,1310	34,0607
9	0,11	15,13	15,4050	0,0756	51,4972
10	0,09	12,21	14,4969	5,2301	101,9323
11	0,08	17,98	14,1564	14,6199	18,7156
12	0,07	13,02	13,4072	0,1500	86,2327
13	0,05	12,04	12,4538	0,1712	105,3939
Σ	3,44	289,98	-	29,8984	1 773,3987

Z předchozí tabulky vyplývá, že jednotlivé součty čtverců jsou

$$S_y = 1773{,}3987\,,$$

$$S_{y,R} = 29{,}8984\,,$$

a teoretický součet čtverců je

$$S_{y,T} = 1773{,}3987 - 29{,}8984 = 1743{,}5003\,.$$

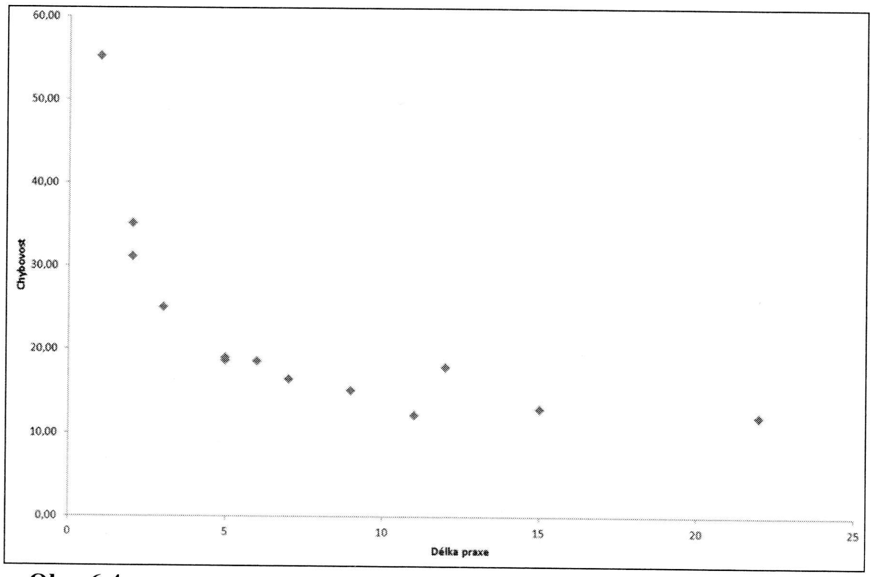
Koeficient determinace určíme jako podíl

$$R^2 = \frac{1743{,}5003}{1773{,}3987} = 1 - \frac{29{,}8984}{1773{,}3987} = 0{,}9831\,.$$

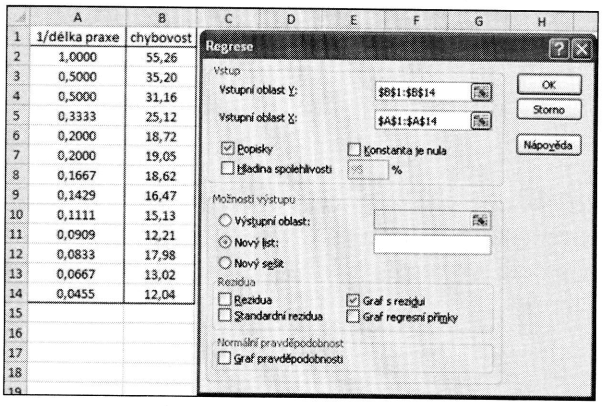
Z hodnoty vypočteného koeficientu determinace vyplývá, že pomocí zvoleného regresního modelu se podařilo vysvětlit 98,31 % variability hodnot proměnné chybovost pracovníka, kde vysvětlující proměnnou je převrácená hodnota délky praxe.

Excel

V grafu na Obr. 6.4 jsou znázorněny hodnoty chybovosti v závislosti na délce praxe. Je patrné, že s růstem praxe klesá chybovost pracovníka. Z uvedeného grafu je patrné, že vhodným modelem by mohla být regresní hyperbola.



Obr. 6.4



	A	B
1	1/délka praxe	chybovost
2	1,0000	55,26
3	0,5000	35,20
4	0,5000	31,16
5	0,3333	25,12
6	0,2000	18,72
7	0,2000	19,05
8	0,1667	18,62
9	0,1429	16,47
10	0,1111	15,13
11	0,0909	12,21
12	0,0833	17,98
13	0,0667	13,02
14	0,0455	12,04

Podle výše uvedeného tvaru regresní hyperboly budeme jako vstupní hodnoty vysvětlující proměnné uvažovat převrácené hodnoty délky praxe, tj. do *Vstupní oblasti X* vložíme dříve připravené převrácené hodnoty.

Pokud modelujeme regresní hyperbolu v Excelu, zvolíme opět nabídky **Analýza dat** a **Regrese**, stejně jako v případě regresní přímky (viz předchozí dialogové okno).

Získaný výstup je opět rozdělen do tří tabulek. V první tabulce jsou uvedeny *Regresní statistiky*, které hodnotí model regresní hyperboly.

VÝSLEDEK	
Regresní statistika	
Násobné R	0,991534492
Hodnota spolehlivosti R	0,98314065
Nastavená hodnota spolehlivosti	0,981607981
Chyba stř. hodnoty	1,648645452
Pozorování	13

Ve druhé tabulce je celkový F-test a hodnoty jednotlivých součtů čtverců.

ANOVA					
	Rozdíl	*SS*	*MS*	*F*	*Významnost F*
Regrese	1	1743,500358	1743,5	641,4569	4,18921E-11
Rezidua	11	29,89835008	2,718032		
Celkem	12	1773,398708			

Třetí tabulka opět obsahuje odhady jednotlivých parametrů regresní hyperboly a dílčí t-testy.

	Koeficienty	*Chyba stř. hodnoty*	*t Stat*	*Hodnota P*	*Dolní 95%*	*Horní 95%*	*Dolní 95,0%*	*Horní 95,0%*
Hranice	10,41058119	0,655498437	15,88193	6,23695E-09	8,967838861	11,853324	8,967838861	11,85332353
1/délka praxe	44,94986223	1,774780357	25,327	4,18921E-11	41,043597	48,856127	41,043597	48,85612746

◘

Příklad 6.3

Využijte data z předchozího příkladu (chybovost případů při řešení obtížných úkolů (v procentech) v závislosti na délce jejich praxe (v měsících)). Pomocí regresní logaritmické funkce vyjádřete závislost chybovosti pracovníka na jeho praxi a zhodnoťte kvalitu regresního modelu.

Řešení:

Tvar regresní logaritmické funkce je možné vyjádřit následujícím způsobem

$$\eta = \beta_0 + \beta_1 \ln x\,.$$

Vzhledem k tomu, že i regresní logaritmická funkce je lineární z hlediska regresních parametrů, je možné její parametry odhadnout pomocí metody nejmenších čtverců. Při odhadování jejích parametrů se využívají uvedené vzorce, přičemž za hodnoty vysvětlující proměnné se dosazují přirozené logaritmy hodnot proměnné chybovost pracovníků, tedy vzorce je možné upravit do tvaru

$$b_1 = \frac{n\sum_{i=1}^{n} y_i \cdot \ln x_i - \sum_{i=1}^{n} \ln x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} (\ln x_i)^2 - (\sum_{i=1}^{n} \ln x_i)^2},$$

resp.

$$b_0 = \frac{\sum_{i=1}^{n} y_i}{n} - b_1 \frac{\sum_{i=1}^{n} \ln x_i}{n}.$$

Pomocné výpočty uspořádejme do tabulky

Tab. 6.7

i	$\ln x_i$	y_i	$\ln x_i y_i$	$(\ln x_i)^2$
1	0,0000	55,26	0,0000	0,0000
2	0,6931	35,2	24,3988	0,4805
3	0,6931	31,16	21,5985	0,4805
4	1,0986	25,12	27,5971	1,2069
5	1,6094	18,72	30,1287	2,5903
6	1,6094	19,05	30,6598	2,5903
7	1,7918	18,62	33,3626	3,2104
8	1,9459	16,47	32,0491	3,7866
9	2,1972	15,13	33,2440	4,8278
10	2,3979	12,21	29,2783	5,7499
11	2,4849	17,98	44,6786	6,1748
12	2,7081	13,02	35,2588	7,3335
13	3,0910	12,04	37,2162	9,5545
Σ	22,3206	289,98	379,4705	47,9859

Dosazením do výše uvedených vzorců získáme odhady parametrů regresní logaritmické funkce:

$$b_1 = \frac{13 \cdot 379{,}4705 - 22{,}3206 \cdot 289{,}98}{13 \cdot 47{,}9859 - 22{,}3206^2} = 43{,}3484\,,$$

$$b_0 = \frac{289{,}98}{13} - 43{,}3484 \cdot \frac{22{,}3206}{13} = -12{,}2555\,.$$

Výsledný tvar regresní logaritmické funkce je tedy možné vyjádřit následovně:

$$Y = -12{,}2555 + 43{,}3484 \cdot \ln x\,.$$

Ke zhodnocení kvality tohoto regresního modelu pomocí koeficientu determinace je opět třeba stanovit další pomocné výpočty, které jsou uspořádány do tabulky

Tab. 6.8

i	$\ln x_i$	y_i	Y_i	$(y_i - Y_i)^2$	$(y_i - \bar{y})^2$
1	0,0000	55,26	43,3484	141,8858	1 085,9560
2	0,6931	35,20	34,8536	0,1200	166,2513
3	0,6931	31,16	34,8536	13,6424	78,3906
4	1,0986	25,12	29,8844	22,6994	7,9177
5	1,6094	18,72	23,6240	24,0490	12,8605
6	1,6094	19,05	23,6240	20,9212	10,6025
7	1,7918	18,62	21,3895	7,6703	13,5877
8	1,9459	16,47	19,5003	9,1830	34,0607
9	2,1972	15,13	16,4204	1,6650	51,4972
10	2,3979	12,21	13,9610	3,0662	101,9323
11	2,4849	17,98	12,8947	25,8605	18,7156
12	2,7081	13,02	10,1599	8,1799	86,2327
13	3,0910	12,04	5,4662	43,2149	105,3939
Σ	22,3206	289,98	-	322,1577	1 773,3987

Z tabulky vyplývá, že jednotlivé součty čtverců jsou

$$S_y = 1773{,}3987\,,$$

$$S_{y,R} = 322{,}1577\,,$$

a teoretický součet čtverců je

$$S_{y,T} = 1773{,}3987 - 322{,}1577 = 1451{,}2410\,.$$

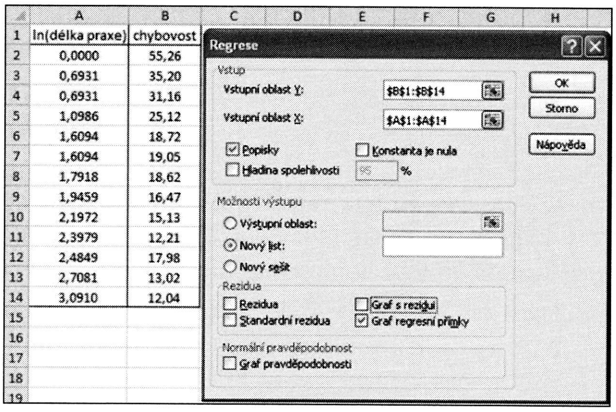
Koeficient determinace určíme jako podíl

$$R^2 = \frac{1451{,}2410}{1773{,}3987} = 1 - \frac{322{,}1577}{1773{,}3987} = 0{,}8183\,.$$

Z hodnoty vypočteného koeficientu determinace vyplývá, že pomocí zvoleného regresního modelu se podařilo vysvětlit 81,83 % variability hodnot proměnné chybovost pracovníka, kde vysvětlující proměnnou je přirozený logaritmus délky praxe.

Excel

Pokud modelujeme regresní logaritmickou funkci v Excelu, postupuje analogicky, jako ve výše uvedených případech, tj. zvolíme opět nabídky **Analýza dat** a **Regrese**. Podle výše uvedeného tvaru regresní logaritmické funkce budeme jako vstupní hodnoty vysvětlující proměnné uvažovat přirozené logaritmy hodnoty délky praxe, tj. do *Vstupní oblasti X* vložíme dříve připravené logaritmované hodnoty vysvětlující proměnné. K logaritmování původních hodnot využijeme funkci *LN*.



	A	B
1	ln(délka praxe)	chybovost
2	0,0000	55,26
3	0,6931	35,20
4	0,6931	31,16
5	1,0986	25,12
6	1,6094	18,72
7	1,6094	19,05
8	1,7918	18,62
9	1,9459	16,47
10	2,1972	15,13
11	2,3979	12,21
12	2,4849	17,98
13	2,7081	13,02
14	3,0910	12,04

Výstup i v tomto případě je rozdělen do tří samostatných tabulek, kde v první z nich jsou *Regresní statistiky*, hodnotící model logaritmické funkce.

VÝSLEDEK	
Regresní statistika	
Násobné R	0,904621
Hodnota spolehlivosti R	0,818339
Nastavená hodnota spolehlivosti R	0,801824
Chyba stř. hodnoty	5,411752
Pozorování	13

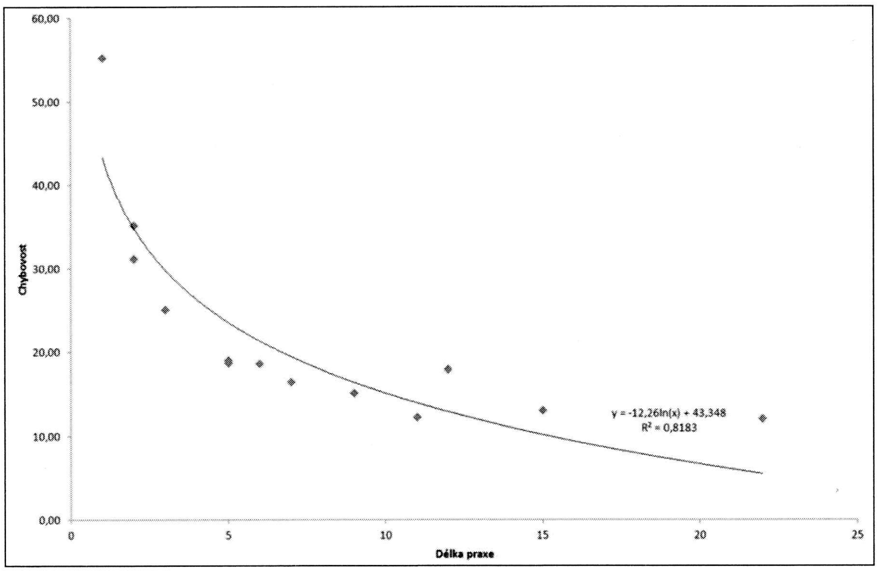
Druhá tabulka obsahuje rozklad součtu čtverců a celkový F-test.

ANOVA					
	Rozdíl	*SS*	*MS*	*F*	*Významnost F*
Regrese	1	1451,241	1451,241	49,55229	2,15624E-05
Rezidua	11	322,1577	29,28706		
Celkem	12	1773,399			

A třetí obsahuje odhady parametrů a dílčí t-testy.

	Koeficienty	*Chyba stř. hodnoty*	*t Stat*	*Hodnota P*	*Dolní 95%*	*Horní 95%*	*Dolní 95,0%*	*Horní 95,0%*
Hranice	43,3484158	3,344906226	12,95953	5,25582E-08	35,98632688	50,71050481	35,98632688	50,71050481
ln(délka praxe)	-12,255484	1,740999356	-7,03934	2,15624E-05	-16,08739784	-8,423570351	-16,08739784	-8,423570351

Pokud bychom postupovali analogicky jako u regresní přímky a v jejím grafu bychom zvolili *Logaritmický trend* získali bychom graf, který kromě napozorovaných hodnot obsahuje i regresní logaritmickou funkci.



Obr. 6.5

◘

Příklad 6.4

Pomocí regresní mocninné funkce vyjádřete závislost zisku firmy (ve stovkách tisíc korun) na počtu jejích poboček. Odhadněte střední hodnotu zisku v případě 5 poboček. Údaje jsou uspořádány do Tab. 6.9.

Tab. 6.9

Počet poboček	1	2	2	3	4	5	6	7	8	9
Zisk	2,20	6,16	8,62	25,13	33,77	57,89	75,06	101,69	132,28	185,72

Řešení:

Ze vstupních údajů, kde počet poboček je vysvětlující proměnná a zisk firmy je vysvětlovaná proměnná, vyplývá, že s rostoucím počtem poboček poroste také zisk firmy. Vzhledem k možným úsporám z rozsahu je také možné předpokládat, že zisk firmy poroste rychleji, než počet jejích poboček. Tomuto předpokladu odpovídá situace zobrazená na Obr. 6.6. Tvar regresní mocninné funkce je možné vyjádřit následujícím způsobem

$$\eta = \beta_0 x^{\beta_1} .$$

Vzhledem k tomu, že regresní mocninná funkce není lineární z hlediska regresních parametrů, je nutné pomocí logaritmické transformace upravit funkci na tvar

$$\ln \eta = \ln \beta_0 + \beta_1 \ln x .$$

Pokud bychom symbolicky rovnici přepsali tak, že logaritmované výrazy označíme hvězdičkou, je možné zapsat

$$\eta^* = \beta_0^* + \beta_1 x^* .$$

Z upraveného výrazu je zřejmé, že aplikací vzorců pro odhady parametrů regresní přímky na transformovaná (zlogaritmovaná) data vysvětlující i vysvětlované proměnné získáme odhady pro regresní mocninnou funkci. Odhad parametru β_0, který získáme standardním postupem, bude však nutné zpětně transformovat (odlogaritmovat), tedy

$$\beta_0 = e^{\beta_0^*} ,$$

kde β_0^* je parametr získaný metodou nejmenších čtverců, kterou jsme aplikovali na výše popsaná zlogaritmovaná data.
Vzorce pro odhad parametrů je možné upravit do tvarů

$$b_1 = \frac{n\sum_{i=1}^{n} \ln y_i \cdot \ln x_i - \sum_{i=1}^{n} \ln x_i \sum_{i=1}^{n} \ln y_i}{n\sum_{i=1}^{n} (\ln x_i)^2 - (\sum_{i=1}^{n} \ln x_i)^2},$$

resp.

$$b_0^* = \frac{\sum_{i=1}^{n} \ln y_i}{n} - b_1 \frac{\sum_{i=1}^{n} \ln x_i}{n}.$$

Pomocné výpočty k odhadu parametrů uspořádejme do Tab. 6.10.

Tab. 6.10

i	$\ln x_i$	$\ln y_i$	$\ln x_i \ln y_i$	$(\ln x_i)^2$
1	0,0000	0,7885	0,0000	0,0000
2	0,6931	1,8181	1,2602	0,4805
3	0,6931	2,1541	1,4931	0,4805
4	1,0986	3,2241	3,5420	1,2069
5	1,3863	3,5196	4,8792	1,9218
6	1,6094	4,0585	6,5320	2,5903
7	1,7918	4,3183	7,7373	3,2104
8	1,9459	4,6219	8,9939	3,7866
9	2,0794	4,8849	10,1579	4,3241
10	2,1972	5,2242	11,4788	4,8278
Σ	13,4950	34,6122	56,0744	22,8288

Dosazením do výše uvedených vzorců získáme odhady parametrů regresní mocninné funkce:

$$b_1 = \frac{10 \cdot 56{,}0744 - 13{,}4950 \cdot 34{,}6122}{10 \cdot 22{,}8288 - 13{,}4950^2} = 2{,}02828,$$

$$b_0^* = \frac{34,6122}{10} - 2,02828 \cdot \frac{13,4950}{10} = 0,72406 .$$

Zpětnou transformací získáme parametr

$$b_0 = e^{0,72406} = 2,06279 .$$

Výsledný tvar odhadnuté regresní mocninné funkce můžeme vyjádřit jako

$$Y = 2,06279 \cdot x^{2,02828} .$$

Ke zhodnocení kvality tohoto regresního modelu pomocí koeficientu determinace je opět třeba stanovit další pomocné výpočty, které jsou uspořádány do Tab. 6.11.

Tab. 6.11

i	$\ln x_i$	$\ln y_i$	$\ln Y_i$	$(\ln y_i - \ln Y_i)^2$	$(\ln y_i - \ln y)^2$
1	0,0000	0,7885	0,7241	0,0041	7,1436
2	0,6931	1,8181	2,1300	0,0973	2,6999
3	0,6931	2,1541	2,1300	0,0006	1,7086
4	1,0986	3,2241	2,9524	0,0738	0,0562
5	1,3863	3,5196	3,5359	0,0003	0,0034
6	1,6094	4,0585	3,9884	0,0049	0,3568
7	1,7918	4,3183	4,3582	0,0016	0,7346
8	1,9459	4,6219	4,6709	0,0024	1,3473
9	2,0794	4,8849	4,9417	0,0032	2,0269
10	2,1972	5,2242	5,1806	0,0019	3,1082
Σ	13,4950	34,6122	-	0,1901	19,1856

Z Tab. 6.11 vyplývá, že jednotlivé součty čtverců jsou následující

$$S_y = 19,1856 ,$$
$$S_{y,R} = 0,1901 ,$$

a teoretický součet je

$$S_{y,T} = 19,1856 - 0,1901 = 18,9955 .$$

Koeficient determinace určíme jako podíl

$$R^2 = \frac{18,9955}{19,1856} = 1 - \frac{0,1901}{19,1856} = 0,99009\,.$$

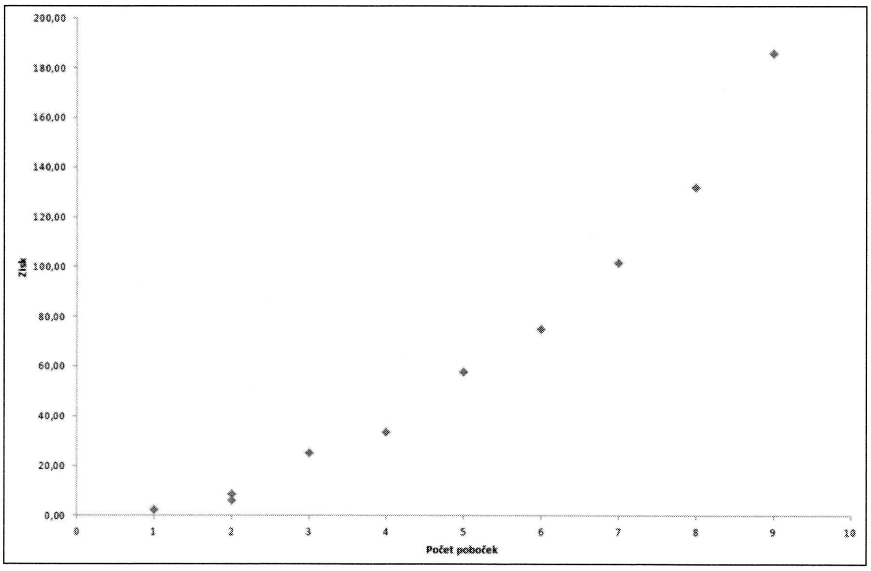
Z hodnoty vypočteného koeficientu determinace vyplývá, že pomocí daného regresního modelu se podařilo vysvětlit 99,009 % variability hodnot proměnné logaritmus zisku firmy.

Proveďme nyní odhad střední hodnoty zisku firmy, která má 5 poboček. Dosadíme číslo 5 do rovnice regresní mocninné funkce,

$$Y = 2,06279 \cdot 5^{2,02828} = 53,9711\,.$$

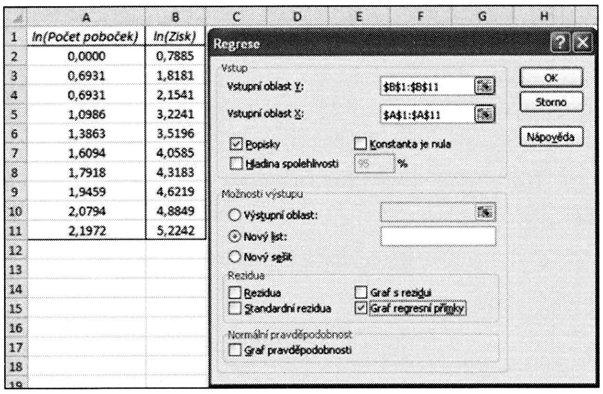
Excel

V grafu na Obr. 6.6 jsou zachyceny hodnoty zisku firmy v závislosti na počtu jejích poboček. Je zřejmé, že zisk firmy roste rychleji než počet jejích poboček, a proto, jak bylo uvedeno výše, by vhodným modelem mohla být regresní mocninná funkce.



Obr. 6.6

V případě, že modelujeme závislost pomocí regresní mocninné funkce, postupujeme opět analogicky jako ve výše uvedených případech, tj. zvolíme nabídky **Analýza dat** a **Regrese**.
Podle výše uvedeného tvaru regresní mocninné funkce budeme jako vstupní hodnoty vysvětlující proměnné uvažovat přirozené logaritmy zisku, tj. do *Vstupní oblasti X* vložíme tyto předem připravené zlogaritmované hodnoty vysvětlující proměnné. Do *Vstupní oblasti Y* vložíme předem připravené zlogaritmované hodnoty vysvětlované proměnné (zlogaritmovaný počet poboček).



Získané výstupy jsou opět rozděleny do tří tabulek.

Charakteristiky, které hodnotí model regresní mocninné funkce, jsou uvedeny v tabulce *Regresní statistiky*.

VÝSLEDEK	
Regresní statistika	
Násobné R	0,995033
Hodnota spolehlivosti R	0,99009
Nastavená hodnota spolehlivosti R	0,988851
Chyba stř. hodnoty	0,154162
Pozorování	10

Celkový F-test a rozklady součtů čtverců jsou v následující tabulce.

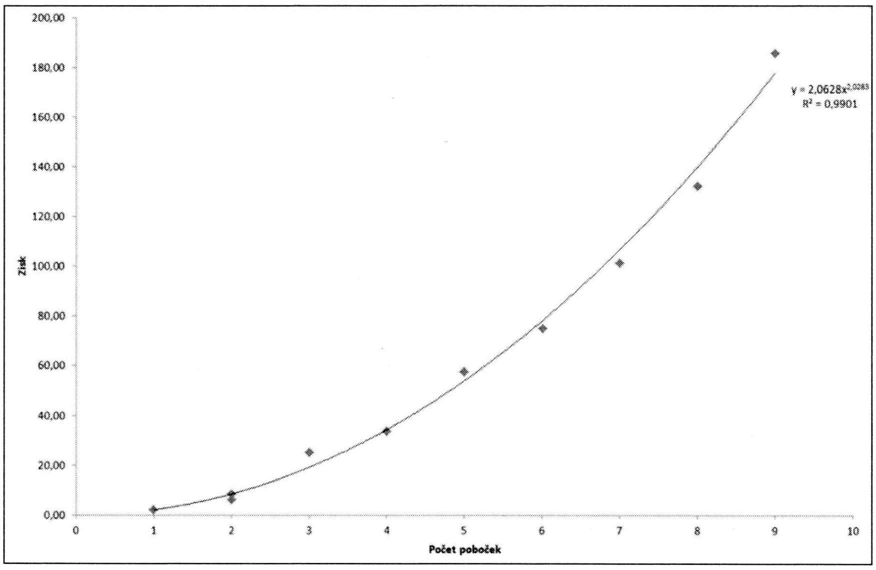
ANOVA					
	Rozdíl	*SS*	*MS*	*F*	*Významnost F*
Regrese	1	18,99547	18,99547	799,2698	2,64772E-09
Rezidua	8	0,190128	0,023766		
Celkem	9	19,1856			

Odhady parametrů regresní mocninné funkce jsou uvedeny v poslední tabulce.

	Koeficienty	*Chyba stř. hodnoty*	*t Stat*	*Hodnota P*	*Dolní 95%*	*Horní 95%*	*Dolní 95,0%*	*Horní 95,0%*
Hranice	0,724058528	0,108398386	6,679606193	0,000155955	0,474091403	0,974025654	0,474091403	0,974025654
ln(Počet poboček)	2,028280325	0,071743289	28,27135948	2,64772E-09	1,862840003	2,193720648	1,862840003	2,193720648

Zde je třeba připomenout, že hodnota parametru v uvedeném výstupu odpovídá výše popsanému parametru, který jsme označili b_0^*, tj. musíme provést zpětnou transformaci odlogaritmováním.

Pokud bychom chtěli získat graf, který obsahuje napozorované hodnoty, formálně zapsaný model regresní mocninné funkce a hodnotu koeficientu determinace, postupovali bychom opět analogicky jako u výše uvedených případů a zvolili bychom *Mocninný trend*. Tím bychom získali následující graf.



Obr. 6.7

Povšimněme si, že tvar rovnice regresní mocninné funkce již obsahuje hodnotu odlogaritmovaného parametru b_0.

□

Příklad 6.5
Využijte data z předchozího příkladu a pomocí regresní exponenciální funkce vyjádřete závislost zisku firmy (ve stovkách tisíc korun) na počtu jejích poboček. Odhadněte střední hodnotu zisku v případě 5 poboček.

Řešení:
Tvar regresní exponenciální funkce je možné vyjádřit následujícím způsobem:

$$\eta = \beta_0 \beta_1^x .$$

Vzhledem k tomu, že regresní exponenciální funkce není lineární z hlediska regresních parametrů, je nutné pomocí logaritmické transformace upravit funkci na tvar

$$\ln \eta = \ln \beta_0 + x \cdot \ln \beta_1 .$$

Pokud bychom opět symbolicky rovnici přepsali tak, že logaritmované výrazy označíme hvězdičkou, je možné přepsat model do tvaru

$$\eta^* = \beta_0^* + x \cdot \beta_1^* .$$

Z upraveného výrazu je zřejmé, že aplikací vzorců pro odhady parametrů regresní přímky na transformovaná (zlogaritmovaná) data vysvětlované proměnné získáme odhady pro regresní exponenciální funkci. Hodnoty vysvětlující proměnné jsou při výpočtech netransformované, na rozdíl od regresní mocninné funkce. Oba získané odhady parametrů, tj. jak odhad parametru β_0, tak v tomto případě i odhad parametru β_1, bude nutné zpětně transformovat (odlogaritmovat) a to následujícím způsobem

$$\beta_0 = e^{\beta_0^*},$$
$$\beta_1 = e^{\beta_1^*},$$

kde β_0^* a β_1^* jsou parametry získané metodou nejmenších čtverců, aplikovanou na výše popsaná zlogaritmovaná data.

Vzorce pro odhad parametrů je možné upravit do tvarů

$$b_1^* = \frac{n\sum_{i=1}^{n} x_i \cdot \ln y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} \ln y_i}{n\sum_{i=1}^{n} (x_i)^2 - (\sum_{i=1}^{n} x_i)^2},$$

resp.

$$b_0^* = \frac{\sum_{i=1}^{n} \ln y_i}{n} - b_1^* \frac{\sum_{i=1}^{n} x_i}{n}.$$

Pomocné výpočty k odhadu parametrů uspořádejme do Tab. 6.12.

Tab. 6.12

i	x_i	$\ln y_i$	$x_i \ln y_i$	$(x_i)^2$
1	1	0,7885	0,7885	1
2	2	1,8181	3,6362	4
3	2	2,1541	4,3082	4
4	3	3,2241	9,6722	9
5	4	3,5196	14,0783	16
6	5	4,0585	20,2927	25
7	6	4,3183	25,9097	36
8	7	4,6219	32,3535	49
9	8	4,8849	39,0794	64
10	9	5,2242	47,0182	81
Σ	47	34,6122	197,1367	289

Dosazením do výše uvedených vzorců získáme odhady parametrů regresní exponenciální funkce

$$b_1^* = \frac{10 \cdot 197{,}1367 - 47 \cdot 34{,}6122}{10 \cdot 289 - 47^2} = 0{,}506013\,,$$

$$b_0^* = \frac{34{,}6122}{10} - 0{,}506013 \cdot \frac{47}{10} = 1{,}08295\,.$$

Zpětnou transformací získáme parametry

$$b_0 = e^{1{,}08295} = 2{,}9534\,,$$

$$b_1 = e^{0{,}506013} = 1{,}6587\,.$$

Výsledný tvar odhadnuté regresní exponenciální funkce můžeme vyjádřit

$$Y = 2{,}9534 \cdot 1{,}6587^x\,.$$

Ke zhodnocení kvality tohoto regresního modelu pomocí koeficientu determinace je opět třeba stanovit další pomocné výpočty, které jsou uspořádány do Tab. 6.13.

Tab. 6.13

i	x_i	$\ln y_i$	$\ln Y_i$	$(\ln y_i - \ln Y_i)^2$	$(\ln y_i - \ln y)^2$
1	1	0,7885	1,58897	0,64082	7,14365
2	2	1,8181	2,09498	0,07668	2,69991
3	2	2,1541	2,09498	0,00349	1,70860
4	3	3,2241	2,60099	0,38821	0,05624
5	4	3,5196	3,10701	0,17021	0,00341
6	5	4,0585	3,61302	0,19849	0,35680
7	6	4,3183	4,11904	0,03970	0,73457
8	7	4,6219	4,62505	0,00001	1,34725
9	8	4,8849	5,13106	0,06059	2,02693
10	9	5,2242	5,63708	0,17043	3,10825
Σ	47	34,61218	-	1,74863	19,18560

Z Tab. 6.13 vyplývá, že jednotlivé součty čtverců jsou následující

$$S_y = 19,1856\,,$$

$$S_{y,R} = 1,74863\,,$$

a teoretický součet je

$$S_{y,T} = 19,1856 - 1,74863 = 17,4369\,.$$

Koeficient determinace určíme jako podíl

$$R^2 = \frac{17,4369}{19,1856} = 1 - \frac{1,74863}{19,1856} = 0,9088\,.$$

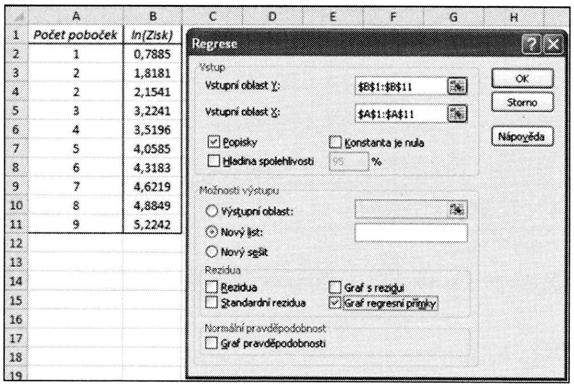
Z hodnoty vypočteného koeficientu determinace vyplývá, že pomocí daného regresního modelu se podařilo vysvětlit 90,88 % variability hodnot proměnné logaritmus zisk firmy.

Proveďme nyní odhad střední hodnoty zisku firmy, která má 5 poboček. Dosadíme číslo 5 do rovnice regresní exponenciální funkce, tj.

$$Y = 2,9534 \cdot 1,6587^5 = 37,082\,.$$

Excel

Pokud modelujeme regresní exponenciální funkci v Excelu, postupujeme opět analogicky volbou **Analýzy dat** a **Regrese**.
Podle výše uvedeného tvaru regresní exponenciální funkce budeme jako vstupní hodnoty vysvětlované proměnné uvažovat přirozené logaritmy hodnoty zisku, tj. do *Vstupní oblasti Y* vložíme tyto předem připravené zlogaritmované hodnoty vysvětlující proměnné. Hodnoty vysvětlující proměnné jsou bez transformace.



Výstupy regresní analýzy jsou opět rozděleny do 3 tabulek.

VÝSLEDEK	
Regresní statistika	
Násobné R	0,95334
Hodnota spolehlivosti R	0,908857
Nastavená hodnota spolehlivosti R	0,897464
Chyba stř. hodnoty	0,467524
Pozorování	10

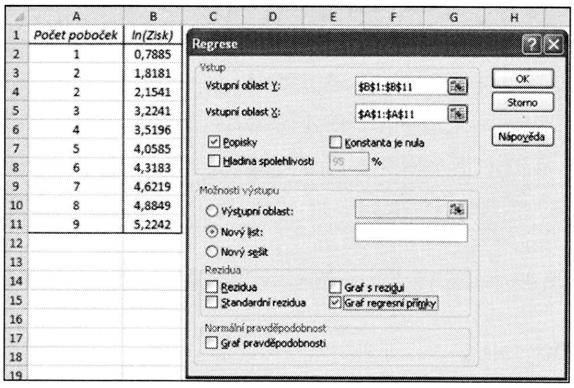
ANOVA					
	Rozdíl	*SS*	*MS*	*F*	*Významnost F*
Regrese	1	17,43697	17,43697	79,77433	1,95988E-05
Rezidua	8	1,74863	0,218579		
Celkem	9	19,1856			

	Koeficienty	*Chyba stř. hodnoty*	*t Stat*	*Hodnota P*	*Dolní 95%*	*Horní 95%*	*Dolní 95,0%*	*Horní 95,0%*
Hranice	1,082954915	0,304564518	3,555749	0,007446734	0,380627878	1,785281952	0,380627878	1,785281952
Počet poboček	0,506013356	0,056653975	8,931648	1,95988E-05	0,375369056	0,636657656	0,375369056	0,636657656

Odhady parametrů, které jsou uvedeny ve sloupci *koeficienty* v rámci třetí tabulky výstupu, jsou opět uvedeny ve zlogaritmovaném tvaru, a proto je třeba provést jejich transformaci odlogaritmováním.

Excel

Pokud modelujeme regresní exponenciální funkci v Excelu, postupujeme opět analogicky volbou **Analýzy dat** a **Regrese**.
Podle výše uvedeného tvaru regresní exponenciální funkce budeme jako vstupní hodnoty vysvětlované proměnné uvažovat přirozené logaritmy hodnoty zisku, tj. do *Vstupní oblasti Y* vložíme tyto předem připravené zlogaritmované hodnoty vysvětlující proměnné. Hodnoty vysvětlující proměnné jsou bez transformace.



Výstupy regresní analýzy jsou opět rozděleny do 3 tabulek.

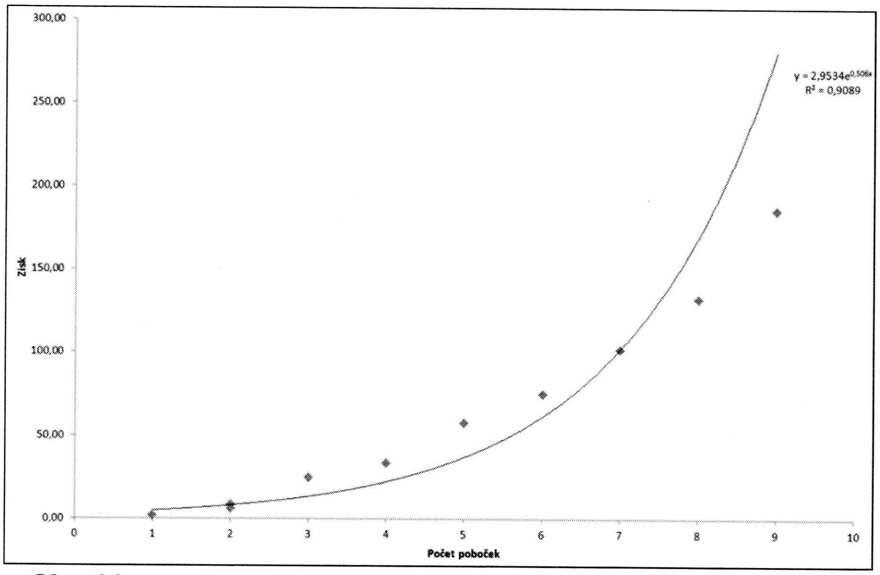
VÝSLEDEK	
Regresní statistika	
Násobné R	0,95334
Hodnota spolehlivosti R	0,908857
Nastavená hodnota spolehlivosti R	0,897464
Chyba stř. hodnoty	0,467524
Pozorování	10

ANOVA					
	Rozdíl	*SS*	*MS*	*F*	*Významnost F*
Regrese	1	17,43697	17,43697	79,77433	1,95988E-05
Rezidua	8	1,74863	0,218579		
Celkem	9	19,1856			

	Koeficienty	*Chyba stř. hodnoty*	*t Stat*	*Hodnota P*	*Dolní 95%*	*Horní 95%*	*Dolní 95,0%*	*Horní 95,0%*
Hranice	1,082954915	0,304564518	3,555749	0,007446734	0,380627878	1,785281952	0,380627878	1,785281952
Počet poboček	0,506013356	0,056653975	8,931648	1,95988E-05	0,375369056	0,636657656	0,375369056	0,636657656

Odhady parametrů, které jsou uvedeny ve sloupci *koeficienty* v rámci třetí tabulky výstupu, jsou opět uvedeny ve zlogaritmovaném tvaru, a proto je třeba provést jejich transformaci odlogaritmováním.

Graf, který obsahuje napozorované hodnoty, model regresní exponenciální funkce a hodnotu koeficientu determinace, získáme volbou *Exponenciální trend* ve výše uvedeném postupu.



Obr. 6.8

Cvičení

1. U 13 studentů byla zjišťována jejich výška (v cm) a jejich hmotnost (v kg). Vstupní údaje jsou uspořádány do tabulky.

   a) Pomocí regresní přímky vyjádřete závislost hmotnosti studenta na jeho výšce.
   b) Pomocí testu ověřte, zda se jedná o vhodný model.
   c) Zhodnoťte kvalitu regresního modelu pomocí vhodného koeficientu.
   d) Vyjádřete intenzitu závislosti mezi hmotností studenta a jeho výškou.
   e) Interpretujte věcně hodnotu vypočteného regresního koeficientu b_1.
   f) Odhadněte střední hodnotu hmotnosti studenta, který měří 180 cm.

Výška	Hmotnost
180	72
197	86
205	81
202	85
180	71
160	63
183	71
185	73
183	72
150	55
170	65
190	75
170	67

2. Uvažujte závislost nabízeného množství výrobku (v kusech), které je ochoten výrobce nabízet při daných cenách (v tisících korun). Vstupní údaje jsou uspořádány do tabulky.

   a) Modelujte závislost nabízeného množství na ceně pomocí vybraných regresních funkcí (regresní přímka, hyperbola, logaritmická funkce, exponenciální a mocninná funkce).
   b) Popište kvalitu všech vybraných funkcí.
   c) Vyberte nejvhodnější model a výběr zdůvodněte.
   d) Určete, jaká je intenzita závislosti nabízeného množství na ceně.

e) Odhadněte střední hodnotu nabízeného množství při ceně 20 tis. Kč.

Cena	Nabízené množství
10	5
11	14
12	16
13	17
14	18
16	21
18	24
19	25
20	26
21	25
22	29
25	33
26	34
28	37
30	39
32	42
34	52

Výsledky

1.

a) $Y = -23{,}3191 + 0{,}526178 \cdot x$.

b) $F = 142{,}00$; p-hodnota = 1,25E-07 ; na 5% i 1% hladině významnosti je prokázána statisticky významná závislost.

c) $R^2 = 0{,}9281$,

d) $r_{yx} = 0{,}9634$,

e) při růstu výšky studenta o 1 centimetr je možné očekávat, že se střední hodnota hmotnosti zvýší o 0,5262 kilogramu,

f) $Y_{180} = 71{,}3929$; $P(69{,}9064 < \eta_{180} < 72{,}8793) = 0{,}95$.

2.

a)

Přímka:	$Y = -4{,}43776 + 1{,}51693 \cdot x$.
Hyperbola:	$Y = 54{,}6868 - 500{,}436/x$.
Log. funkce:	$Y = -60{,}3456 + 29{,}4627 \cdot \ln x$.
Exp. funkce:	$Y = \exp(1{,}85848 + 0{,}0638816 \cdot x) = 6{,}41398 \cdot 1{,}065966^x$.
Mocninná fce:	$Y = 0{,}49852 \cdot x^{1{,}30836}$.

b) Přímka: $R^2 = 0{,}9595$.
Hyperbola: $R^2 = 0{,}8746$.
Log. funkce: $R^2 = 0{,}9364$.
Exp. funkce: $R^2 = 0{,}7872$.
Mocninná fce: $R^2 = 0{,}8543$.

c) Nejvhodnější model je regresní přímka, a to dle grafu napozorovaných hodnot, nejvyšší hodnoty R^2; celkový F-test vychází významný (na 5% i 1% hladině významnosti) – F = 355,0 , u obou dílčích t-testů je zamítnuta testovaná hypotéza o nulové hodnotě příslušného parametru (na 5% hladině významnosti).

d) $r_{yx} = 0{,}9795$ (regresní přímka), jedná se velmi silnou lineární závislost (s růstem ceny roste i nabízené množství).

e) $Y_{20} = 25{,}90$; $P(24{,}6377 < \eta_{20} < 27{,}1639) = 0{,}95$.

6.2 Regresní parabola

Příklad 6.6

U 8 automobilů byly měřeny roční náklady na údržbu (v korunách) v závislosti na jejich stáří (v letech). Pomocí regresní paraboly popište závislost ročních nákladů na stáří automobilu. Dále zhodnoťte kvalitu tohoto regresního modelu a odhadněte střední hodnotu nákladů na údržbu automobilu, který je starý 5 let. Vstupní údaje jsou uspořádány do Tab. 6.14.

Tab. 6.14

Stáří	1	2	3	4	5	6	7	8
Náklady	1 200	1 140	1 100	1 100	1 220	1 260	1 420	1 680

Řešení:

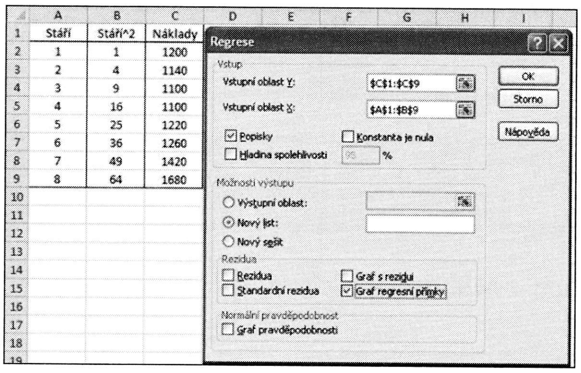
Tvar regresní paraboly je možné vyjádřit následujícím způsobem:

$$\eta = \beta_0 + \beta_1 x + \beta_2 x^2 .$$

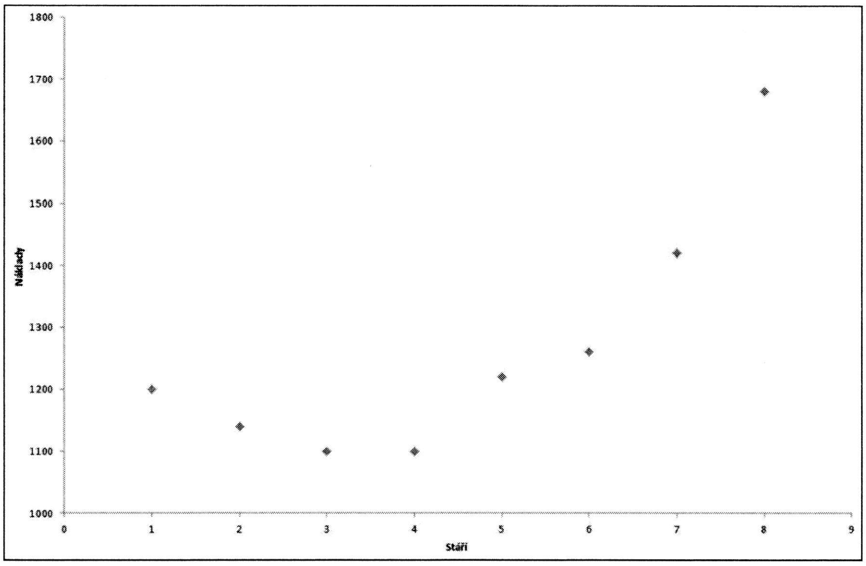
Postup řešení regresní paraboly si ukážeme pro jednoduchost pouze s využitím Excelu.

Excel

Pokud modelujeme regresní parabolu v Excelu, postupujeme opět analogicky volbou nabídek **Analýza dat** a **Regrese**. Do *Vstupní oblasti X* musíme vložit současně hodnoty stáří vozu a jeho druhé mocniny, kterou si musíme připravit předem.



Z grafu napozorovaných hodnot, viz Obr. 6.9, je patrné, že průběh závislosti nákladů na stáří automobilu bude zřejmě vhodné modelovat pomocí regresní paraboly.



Obr. 6.9

Výstupy získané z Excelu jsou rozděleny opět do tří tabulek:

VÝSLEDEK	
Regresní statistika	
Násobné R	0,991826
Hodnota spolehlivosti R	0,983719
Nastavená hodnota spolehlivosti R	0,977207
Chyba stř. hodnoty	29,79294
Pozorování	8

ANOVA					
	Rozdíl	*SS*	*MS*	*F*	*Významnost F*
Regrese	2	268161,9	134081	151,0569	3,38203E-05
Rezidua	5	4438,095	887,619		
Celkem	7	272600			

	Koeficienty	*Chyba stř. hodnoty*	*t Stat*	*Hodnota P*	*Dolní 95%*	*Horní 95%*	*Dolní 95,0%*	*Horní 95,0%*
Hranice	1338,57143	41,56545531	32,20394	5,42E-07	1231,724024	1445,418833	1231,724024	1445,418833
Stáří	-152,619048	21,19181376	-7,20179	0,000804	-207,0943391	-98,1437561	-207,0943391	-98,1437561
Stáří^2	24,047619	2,298574931	10,46197	0,000138	18,13894408	29,95629401	18,13894408	29,95629401

Z uvedeného výstupu z MS Excel vyplývá, že rovnici regresní paraboly můžeme vyjádřit následovně:

$$Y = 1338,57 - 152,619x + 24,0476x^2 .$$

Vzhledem k tomu, že uvažujeme model, který má oproti doposud výše popsaným modelům vyšší počet parametrů, je pro případné srovnání s uvedenými funkcemi nutné stanovit upravený koeficient determinace, a to podle výše uvedeného vzorce, tedy

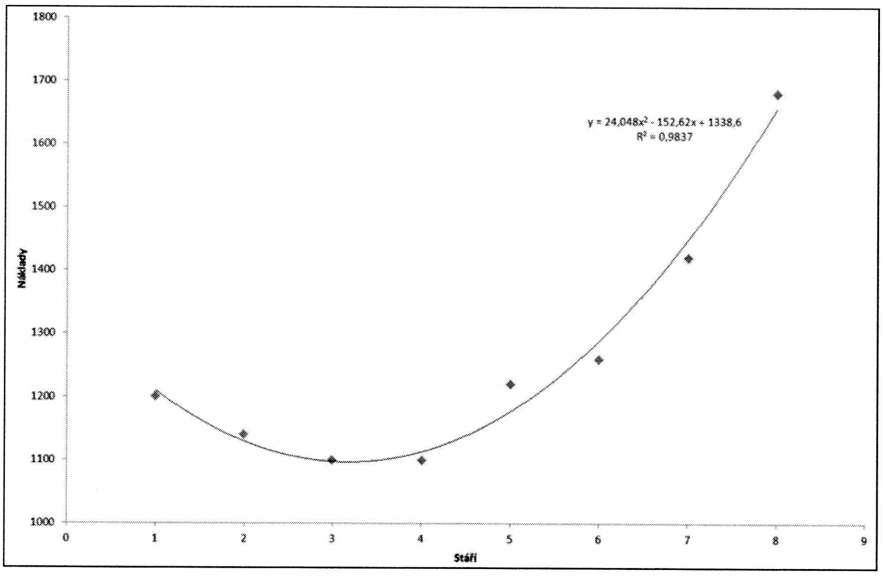
$$R^2_{adj.} = 1 - (1 - 0{,}983719)\frac{8-1}{8-3} = 0{,}977207\,.$$

Tato hodnota je uvedena v první tabulce výstupu a je označena jako *Nastavená hodnota spolehlivosti R*.

Proveďme nyní odhad střední hodnoty nákladů na údržbu automobilu, který je starý 5 let. Získáme jej tak, že dosadíme číslo 5 do rovnice regresní paraboly, tedy

$$Y = 1338{,}57 - 152{,}619 \cdot 5 + 24{,}0476 \cdot 5^2 = 1176{,}67\,.$$

Graf, který obsahuje napozorované hodnoty, rovnici regresní paraboly a hodnotu koeficientu determinace, získáme volbou *Polynomický trend.* Hodnotu *Pořadí* ponecháme na přednastavené hodnotě 2. Pokud bychom chtěli polynom vyššího stupně, museli bychom tuto skutečnost vyjádřit právě nastavením vyšší hodnoty.



Obr. 6.10

□

Cvičení

1. V tabulce jsou uvedeny zisky firmy (v korunách) a počty zákazníků.

   a) Modelujte závislost zisku firmy na počtu zákazníků pomocí regresní paraboly.
   b) Zhodnoťte vhodnost lineárního a kvadratického členu v modelu.
   c) Popište kvalitu daného regresního modelu.
   d) Odhadněte střední hodnotu zisku firmy s 10 zákazníky.

Počet zákazníků	Zisk
1	500
2	746
2	760
3	810
5	917
6	961
6	980
8	1028
9	1052
10	1070
12	1084
13	1082
17	1060
21	800

Výsledky

1.

   a) $Y = 528{,}119 + 94{,}7928 \cdot x - 3{,}88085 \cdot x^2$.
   b) Na 5% i 1% hladině významnosti je zamítnuta testovaná hypotéza u obou regresních parametrů.
   c) $R^2 = 0{,}936519$, tj. pomocí daného regresního modelu se podařilo vysvětlit 93,6519 % variability hodnot závisle proměnné (zisk).
   d) $Y_{10} = 1087{,}96$; $P(1046{,}63 < \eta_{20} < 1129{,}29) = 0{,}95$.

6.3 Vícenásobná regrese

Příklad 6.7

Prozkoumejte závislost ceny automobilu (v tisících korun) na počtu ujetých kilometrů (v tisících) a na jeho stáří (v měsících). Vstupní údaje jsou uspořádány do Tab. 6.15.

a) Modelujte závislost pomocí lineární regresní funkce a pomocí vhodného testu zhodnoťte model jako celek.
b) Zhodnoťte vhodnost jednotlivých vysvětlujících proměnných v modelu pomocí vhodných testů.
c) Popište kvalitu daného regresního modelu.
d) Interpretujte věcně hodnoty dílčích regresních parametrů.
e) Odhadněte střední hodnotu ceny automobilu, který je starý 3 roky a má najeto 42 tis. kilometrů.

Tab. 6.15

Cena	Stáří	Počet km
262	42	26
85	60	125
110	57	122
36	93	54
409	9	7
275	37	46
250	11	19
216	53	25
160	44	74
324	26	38
295	24	29
54	85	92
34	95	75
430	6	3
150	62	93
105	73	91
292	35	26
340	37	12
30	95	50
108	72	94

Řešení:
Uvažujme lineární závislost ceny automobilu na počtu ujetých kilometrů a stáří automobilu. Regresní funkci pak můžeme vyjádřit ve tvaru

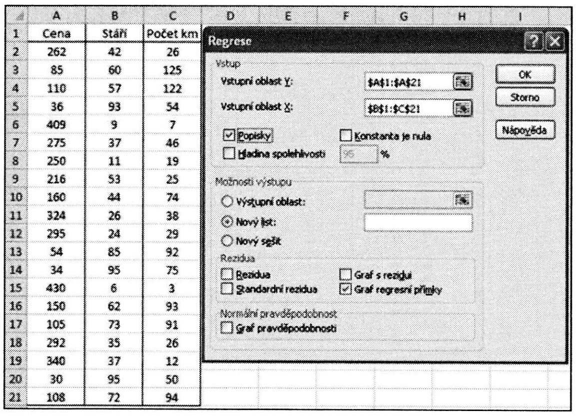
$$\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 ,$$

kde proměnná X_1 vyjadřuje stáří automobilu a proměnná X_2 vyjadřuje počet ujetých kilometrů (v tisících). Vzhledem k tomu, že jde o zcela lineární model, k odhadu parametrů lze využít metodu nejmenších čtverců. Je možné očekávat, že cena automobilu bude klesat jednak s růstem počtu ujetých kilometrů, jednak s růstem jeho stáří.

Výpočet regresní roviny si pro jednoduchost opět ukážeme pouze s využitím Excelu.

Excel

Pokud modelujeme regresní rovinu v Excelu, do *Vstupní oblasti X* musíme vložit současně hodnoty obou vysvětlujících proměnných, tj. stáří automobilu i počet ujetých kilometrů.



Odhady regresních parametrů, celkový F-test, dílčí t-testy i další výsledky jsou patrné z následujícího výstupu z Excelu.

VÝSLEDEK	
Regresní statistika	
Násobné R	0,960113
Hodnota spolehlivosti R	0,921817
Nastavená hodnota spolehlivosti R	0,91262
Chyba stř. hodnoty	37,68454
Pozorování	20

ANOVA					
	Rozdíl	*SS*	*MS*	*F*	*Významnost F*
Regrese	2	284649,6	142324,8	100,2199	3,90331E-10
Rezidua	17	24142,12	1420,125		
Celkem	19	308791,8			

	Koeficienty	*Chyba stř. hodnoty*	*t Stat*	*Hodnota P*	*Dolní 95%*	*Horní 95%*	*Dolní 95,0%*	*Horní 95,0%*
Hranice	421,6552	17,89320905	23,5651	2,019E-14	383,9038256	459,406568	383,9038256	459,406568
Stáří	-3,28191	0,392051903	-8,3711	1,958E-07	-4,109062264	-2,454747842	-4,109062264	-2,454747842
Počet km	-1,02969	0,293674324	-3,50623	0,0027072	-1,649288402	-0,410091075	-1,649288402	-0,410091075

a) Z uvedeného výstupu vyplývá, že rovnici regresní roviny, která popisuje vztah mezi cenou automobilu, jeho stářím a počtem ujetých kilometrů, je možné vyjádřit jako $Y = 421{,}655 - 3{,}28191 \cdot x_1 - 1{,}02969 \cdot x_2$; testovaná hypotéza celkového F-testu je na 5% hladině významnosti je zamítnuta, protože uvedená p-hodnota je 3,903E-10, což znamená, že je menší než zvolená hladina významnosti.

b) Vzhledem k tomu, že p-hodnoty dílčích t-testů u obou vysvětlujících proměnných jsou v porovnání se zvolenou hladinou významnosti (0,05 i 0,01) menší, je možné testované hypotézy o nulových hodnotách regresních parametrů na dané hladině významnosti zamítnout a tím pádem jsme prokázali, že jednak proměnná stáří vozu, jednak proměnná počet ujetých kilometrů je v modelu opodstatněná.

c) Kvalitu zvoleného regresního modelu zhodnotíme pomocí koeficientu determinace $R^2 = 0{,}921817$, ze kterého vyplývá, že pomocí daného modelu se podařilo vysvětlit 92,1817 % variability proměnné cena automobilu. Modifikovaný koeficient determinace pro případ srovnání s modely, které mají jiný počet parametrů, je $R^2_{adj.} = 0{,}91262$.

d) Hodnota parametru $b_1 = -3{,}28191$; udává, že s růstem stáří o jeden měsíc můžeme očekávat, že střední hodnota ceny automobilu poklesne o částku 3281,91 Kč; hodnota parametru $b_2 = -1{,}02969$ nám říká, že s každým dalším najetým tisícem kilometrů poklesne střední hodnota ceny automobilu o 1029,69 Kč.

e) Střední hodnota ceny tři roky starého automobilu, který najel 42 000 km, je $Y_{3;42} = 260\,259$ Kč.

▫

Cvičení

1. Prozkoumejte závislost mzdy pracovníka (v tisících korun) na jeho počtu odpracovaných hodin a počtu úspěšně dokončených výrobků. Vstupní údaje jsou uspořádány do tabulky.
   a) Modelujte závislost pomocí lineární regresní funkce a pomocí vhodného testu zhodnoťte model jako celek.
   b) Zhodnoťte vhodnost jednotlivých vysvětlujících proměnných v modelu pomocí vhodných testů.

Mzda	Počet výrobků	Počet hodin
31	105	180
40	99	196
35	126	180
30	102	172
19	120	140
20	108	148
25	120	160
30	114	164
35	96	180
25	120	160
19	120	140
20	108	148
25	120	160

Výsledky

1.
   a) $Y = -42{,}2037 + 0{,}0349378 \cdot x_1 + 0{,}400239 \cdot x_2$. Na 5% (i na 1%) hladině významnosti byla zamítnuta testovaná hypotéza použitím celkového F-testu, tj. prokázali jsme, že existuje alespoň jedna vysvětlující proměnná, která je v modelu statisticky významná, tj. statisticky významně ovlivňuje mzdu pracovníka. K tomuto výsledku jsme dospěli jednak na základě stanovené p-hodnoty, která je menší než zvolená hladina významnosti (0,05 i 0,01), resp. podle hodnoty testového kritéria F = 116,98, kterou porovnáme s kritickou hodnotou $F_{0,95}(2;10) = 4{,}103$.

b) Vhodnost jednotlivých vysvětlujících proměnných ověříme pomocí dílčích t-testů. *P*-hodnota dílčího t-testu u proměnné počet dokončených výrobků (X_1) vychází 0,5196, což znamená, že na všech obvyklých hladinách významnosti (0,01, 0,05 či 0,10) nejsme oprávněni zamítnout testovanou hypotézu, a tedy na dané hladině významnosti můžeme daný regresní parametr (β_1) považovat za nulový. Z tohoto důvodu je proměnná X_1 v modelu neopodstatněná, a můžeme ji tedy z modelu vyřadit. *P*-hodnota u proměnné počet odpracovaných hodin (X_2) vychází velmi malé číslo, což znamená, že na všech obvyklých hladinách významnosti jsme oprávněni zamítnout testovanou hypotézu o nulové hodnotě regresního parametru (β_2), a tedy tato vysvětlující proměnná je v daném modelu významná.

Jako vhodný model by tedy mohl být zvolen model regresní přímky (po vyjmutí proměnné X_1 z modelu) a výsledný tvar rovnice by byl následující: $Y = -36{,}6957 + 0{,}390528 \cdot x_2$. Hodnota koeficientu determinace je $R^2 = 0{,}957185$, což znamená, že pomocí regresní přímky se podařilo vysvětlit 95,7185 % variability proměnné mzda pracovníka, což představuje převážnou část a daný model můžeme považovat za vhodný.

6.4 Korelační analýza

Příklad 6.8

Pomocí korelačního koeficientu vyjádřete míru lineární závislosti mezi příjmy (X) a výdaji (Y) domácností. Hodnotu korelačního koeficientu interpretujte. Vstupní údaje jsou uspořádány do Tab. 6.16.

Tab. 6.16

Příjem (v tis. Kč)	Výdaj (v tis. Kč)
20	18
25	22
17	15
18	16
19	16
26	22
28	24
28	24
32	29
31	26
26	22
26	22
29	25

Řešení:

Vzorec pro výpočet výběrového korelačního koeficientu mezi proměnnými X a Y můžeme vyjádřit následujícím způsobem

$$r_{yx} = r_{xy} = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sqrt{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}\sqrt{n\sum_{i=1}^{n} y_i^2 - (\sum_{i=1}^{n} y_i)^2}}.$$

Tento koeficient nabývá hodnot z intervalu od – 1 do 1. Záporné hodnoty znamenají nepřímou lineární závislost, kladné hodnoty přímou lineární závislost a hodnota 0 znamená lineární nezávislost.

Pomocné výpočty jsou uspořádány do Tab. 6.17. Hodnotu korelačního koeficientu získáme dosazením do výše uvedeného vzorce, tedy

$$r_{yx} = \frac{13 \cdot 7276 - 325 \cdot 281}{\sqrt{13 \cdot 8421 - 325^2}\,\sqrt{13 \cdot 6291 - 281^2}} = 0{,}990196\,.$$

Z hodnoty vypočteného výběrového korelačního koeficientu vyplývá, že mezi výdajem domácnosti a jejím příjmem existuje velmi silný, přímo úměrný vztah (s růstem příjmu domácnosti roste i její výdaj).

Tab. 6.17

i	x_i	y_i	$x_i y_i$	x_i^2	y_i^2
1	20	18	360	400	324
2	25	22	550	625	484
3	17	15	255	289	225
4	18	16	288	324	256
5	19	16	304	361	256
6	26	22	572	676	484
7	28	24	672	784	576
8	28	24	672	784	576
9	32	29	928	1 024	841
10	31	26	806	961	676
11	26	22	572	676	484
12	26	22	572	676	484
13	29	25	725	841	625
Σ	325	281	7 276	8 421	6 291

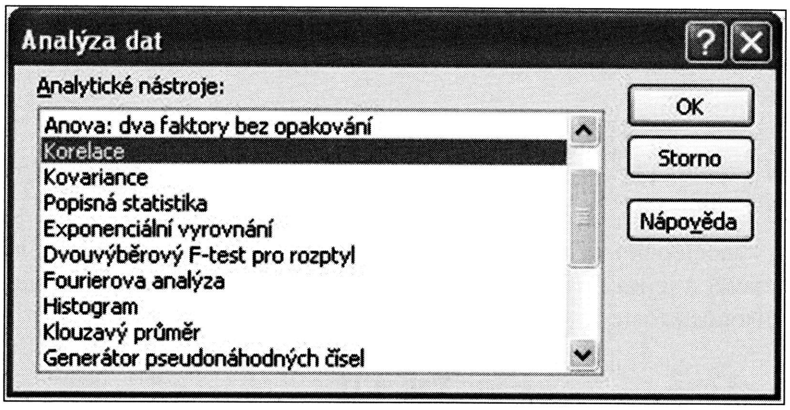
Excel

Pokud chceme získat hodnoty korelačních koeficientů v Excelu, postupujeme následujícím způsobem.

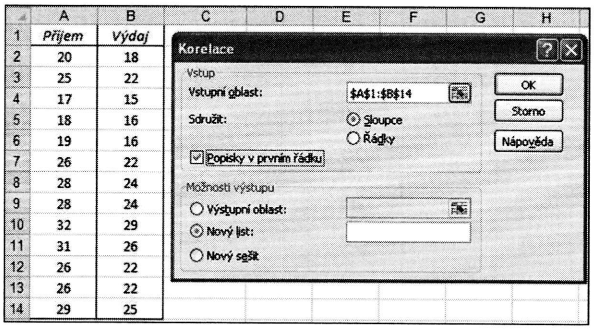
Data

Analýza dat

Korelace



Do vstupního okna vyznačíme *Vstupní oblast*, která obsahuje hodnoty proměnných a v případě, že první řádek obsahuje názvy proměnných, tuto skutečnost také vyznačíme.



Výsledná hodnota korelačního koeficientu je uspořádána do čtvercové symetrické korelační matice, která má následující podobu

	Příjem	*Výdaj*
Příjem	1	
Výdaj	0,990196	1

□

Příklad 6.9

Pomocí korelačního koeficientu vyjádřete závislost poptávaného množství výrobku (v kusech), které je spotřebitel ochoten nakupovat při daných cenách (v korunách). Hodnotu korelačního koeficientu interpretujte a pomocí vhodného testu ověřte, zda se jedná o statisticky významnou závislost. Vstupní údaje jsou uspořádány do Tab. 6.18.

Tab. 6.18

Poptávané množství	Cena
45	40
55	63
38	80
40	89
40	81
55	61
60	50
60	75
75	13
65	36
55	70
55	90
61	80

Řešení:

Pro ověření, zda závislost mezi proměnnými X a Y můžeme považovat na zvolené hladině významnosti za statisticky významnou, použijeme test, kde ověřujeme hypotézu, že hodnota populačního korelačního koeficientu mezi danou dvojicí proměnných je rovna 0, tedy že proměnné jsou lineárně nezávislé (nekorelované)

$$H_0\colon \rho_{yx} = 0,$$

$$H_1\colon \rho_{yx} \neq 0.$$

Testovým kritériem tohoto testu je statistika t, která se stanoví podle vzorce

$$t = \frac{r_{yx}\sqrt{n-2}}{\sqrt{1-r_{yx}^2}}.$$

Předpokládejme, že dvojice hodnot x a y jsou výběrem z dvourozměrného normálního rozdělení. Kritický obor je pak dán nerovností $W_\alpha = \{t; |t| \geq t_{1-\alpha/2}\}$, kde $t_{1\text{-}\alpha/2}$ představuje kvantil t-rozdělení s $(n-2)$ stupni volnosti. Pro velké rozsahy výběru (stačí když $n - 2 > 30$), lze kvantil $t_{1\text{-}\alpha/2}$ nahradit kvantilem $u_{1\text{-}\alpha/2}$ normovaného normálního rozdělení.

Hodnotu korelačního koeficientu získáme postupem, který byl podrobně popsán v předchozím příkladu. Tím získáme výsledek $r_{yx} = -0{,}6248$, což vyjadřuje, že mezi poptávaným množstvím a cenou výrobku je středně silná, nepřímo úměrná závislost (s růstem ceny výrobku dochází k poklesu poptávaného množství).
Ověření statistické významnosti dané závislosti provedeme pomocí testu

$$H_0\text{: } \rho_{yx} = 0\,,$$
$$H_1\text{: } \rho_{yx} \neq 0\,.$$

Testové kritérium stanovíme podle výše uvedeného vzorce jako

$$t = \frac{-0{,}6248}{\sqrt{1-(-0{,}6248)^2}}\sqrt{13-2} = -2{,}65418\,.$$

Kritická hodnota je dána kvantilem $t_{0,975}(11) = 2{,}20$. Vzhledem k tomu, že hodnota testového kritéria spadá do kritického oboru, na 5% hladině významnosti zamítáme testovanou hypotézu a prokázali jsme, že mezi poptávaným množstvím a cenou výrobku existuje statisticky významná závislost.

□

Příklad 6.10
Pomocí vhodného koeficientu změřte těsnost závislosti mezi hodnocením 10 firem, které pochází od dvou nezávislých skupin hodnotitelů. Pořadí firem, sestavená na základě těchto hodnocení, jsou upořádána do Tab. 6.19.

Tab. 6.19

Firma	Pořadí od skupiny 1	Pořadí od skupiny 2
1	2	2
2	1	3
3	3	1
4	5	6
5	6	5
6	9	9
7	8	8
8	7	7
9	4	4
10	10	10

Řešení:
Vzhledem k tomu, že je třeba prozkoumat těsnost závislosti mezi dvěma pořadovými proměnnými, použijeme k jejímu vyjádření Spearmanův korelační koeficient, který se počítá podle vzorce

$$r_S = 1 - \frac{6\sum_{i=1}^{n}(i_x - i_y)^2}{n(n^2-1)},$$

kde i_x a i_y jsou hodnoty pořadových proměnných.
Ověření statistické významnosti dané závislosti provedeme pomocí testu:

$$H_0: \rho_S = 0,$$
$$H_1: \rho_S \neq 0.$$

Testovým kritériem tohoto testu je statistika t, která se stanoví podle vzorce

$$t = \frac{r_S}{\sqrt{1-r_S^2}}\sqrt{n-2}\,.$$

Kritický obor je dán nerovností $W_\alpha = \{t; |t| \geq t_{1-\alpha/2}\}$, kde $t_{1-\alpha/2}$ představuje kvantil t-rozdělení s $(n-2)$ stupni volnosti.
Pomocné údaje uspořádejme do Tab. 6.20, kde i_x a i_y jsou pořadí, která stanovila příslušná skupina hodnotitelů.

Tab. 6.20

Firma	i_x	i_y	$i_x - i_y$	$(i_x - i_y)^2$
1	2	2	0	0
2	1	3	-2	4
3	3	1	2	4
4	5	6	-1	1
5	6	5	1	1
6	9	9	0	0
7	8	8	0	0
8	7	7	0	0
9	4	4	0	0
10	10	10	0	0

Dosazením do výše uvedeného vzorce získáme hodnotu Spearmanova korelačního koeficientu

$$r_S = 1 - \frac{6 \cdot 10}{10(10^2 - 1)} = 0,93939.$$

Z hodnoty vypočteného korelačního koeficientu vyplývá, že mezi pořadím přiřazeným firmám oběma skupinami hodnotitelů existuje velmi silná přímo úměrná korelace.

Nyní otestujme významnost závislosti pomocí testu:

H_0: $\rho_S = 0$,
H_1: $\rho_S \neq 0$.

Dosazením do výše uvedeného vzorce získáme hodnotu testového kritéria t

$$t = \frac{0,93939}{\sqrt{1 - 0,93939^2}} \sqrt{10 - 2} = 7,75.$$

Kritická hodnota je dána kvantilem $t_{0,975}(8) = 2,306$.

Vzhledem k tomu, že hodnota testového kritéria spadá do kritického oboru, na 5% hladině významnosti zamítáme testovanou hypotézu, a tedy jsme prokázali, že mezi hodnocením obou skupin existuje statisticky významná závislost.

◘

Cvičení

1. U 16 směsí byl sledován výskyt jednotlivých látek (proměnné $X_1 - X_3$).
   a) Pomocí vhodných koeficientů stanovte, zda-li se výskyt jednotlivých látek vzájemně ovlivňuje.
   b) Ověřte, zda-li danou závislost je možné považovat za statisticky významnou.

X_1	X_2	X_3
1,2	2,56	4,27
1,7	3,76	2,34
1,9	3,92	1,84
2,2	4,42	1,29
2,4	4,32	1,01
2,6	4,48	0,8
2,7	4,56	0,71
2,9	4,72	0,56
3,1	4,88	0,44
3,3	5,04	0,34
3,2	4,72	0,39
4,1	5,68	0,13
4,2	6,11	0,12
4,4	5,92	0,09
4,6	6,08	0,07
4,8	9,16	0,06

2. Dvě skupiny porotců hodnotily 20 filmů. Na jejich základě bylo stanovena tabulka pořadí jednotlivých filmů.
   a) Pomocí vhodného koeficientu vyjádřete shodu pořadí hodnocení oběma skupinami.
   b) Pomocí vhodného testu ověřte, zda-li je možné danou shodu považovat za statisticky významnou.

Film	Porota 1	Porota 2
1	9	7
2	8	9
3	2	2
4	1	1
5	7	9
6	6	3
7	3	6
8	5	4

Film	Porota 1	Porota 2
9	4	5
10	10	10
11	11	11
12	16	16
13	17	17
14	12	12
15	13	13
16	20	20
17	17	18
18	18	15
19	14	14
20	15	17

3. Modelujte závislost proměnné Y na proměnných X_1 a X_2. Vstupní údaje jsou uspořádány do tabulky.
   a) Vyjádřete rovnici regresní funkce ve tvaru $\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2$.
   b) Zhodnoťte pomocí koeficientu determinace model.
   c) Pomocí vhodných testů ověřte významnost tohoto modelu jako celku a významnost jednotlivých vysvětlujících proměnných.
   d) Pomocí korelačních koeficientů zhodnoťte intenzitu závislosti mezi všemi dvojicemi proměnných.
   e) Navrhněte alternativní model.

Y	X_1	X_2
1008	101	1,6
1146	115	3,9
10	46	1,6
2282	229	7,7
2800	203	9,3
1684	169	5,7
2052	206	7,1
757	76	2,6
886	89	3,3
780	75	2,4

Výsledky

1.

a) $r_{x_1x_2} = 0,8886$, $r_{x_1x_3} = -0,8379$, $r_{x_2x_3} = -0,7238$.

b) Na 1% i 5% hladině významnosti je možné zamítnout testovanou hypotézu o nulové hodnotě korelačního koeficientu u všech dvojic proměnných.

2.

a) $r_S = 0,9616$.

b) Na 1 % i 5% hladině významnosti je zamítnuta testovaná hypotéza o nulové hodnotě Spearmanova korelačního koeficientu.

3.

a) $Y = -170,288 + 5,94106 \cdot x_1 + 162,191 \cdot x_2$.

b) $R^2 = 0,943732$ ($R^2_{upr.} = 0,927656$).

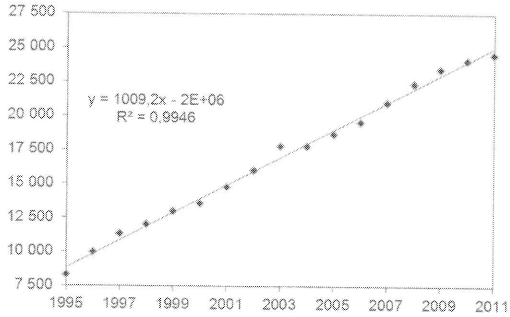
c) P-hodnota u celkového F-testu je menší než zvolená hladina významnosti (0,05 i 0,01), a tak je na dané hladině významnosti možné zamítnout testovanou hypotézu, tj. existuje alespoň jedna vysvětlující proměnná, která je opodstatněná; p-hodnoty u všech dílčích t-testů jsou jsou větší než hladina významnosti (0,05 i 0,01), a tedy nejsme oprávněni na dané hladině významnosti zamítnout testované hypotézy dílčích t-testů.

d) $r_{yx_1} = 0,9546$, $r_{yx_2} = 0,9588$, $r_{x_1x_2} = 0,9399$ – z uvedených hodnot korelačních koeficientů je zřejmé, že mezi všemi dvojicemi hodnot existuje velmi silná přímá lineární závislost.

e) Jak vyplývá z řešení v úloze c) výsledek celkového F-testu a dílčích t-testů je v rozporu, což je zřejmě způsobeno tzv. multikolinearitou, která byla prokázána pomocí korelačního koeficientu $r_{x_1x_2}$, neboť pro danou dvojici je větší (v absolutní hodnotě) než 0,8. Tyto dvě vysvětlující proměnné nemohou být ponechány v modelu současně, a proto z modelu vyřadíme tu proměnnou, která má „nižší“ vazbu na vysvětlovanou proměnnou Y. V našem případě se jedná o proměnnou X_1 (měřeno korelačním koeficientem mezi proměnnými Y a X_1 resp. Y a X_2). Výsledným modelem tedy bude regresní přímka ve tvaru $Y = 9,96797 + 294,365 \cdot x_2$, hodnota $R^2 = 0,919278$.

KAPITOLA VII

ČASOVÉ ŘADY



7

7.1

Přík
V ča
čete
lutní

rok	
poče rozvo	

Zdroj

Řeše
Abso

a ted

Relat

a tedy

Koef

7 Časové řady

7.1 Jednoduché míry dynamiky časové řady

Příklad 7.1
V časové řadě v tabulce jsou uvedeny počty rozvodů v České republice za 12 let. Určete pro tuto řadu absolutní a relativní přírůstky, koeficienty růstu, průměrný absolutní přírůstek a průměrný koeficient růstu.

rok	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
počet rozvodů	29704	31586	31758	32824	33060	31288	31415	31129	31300	29133	30783	28113

Zdroj: http://www.czso.cz/

Řešení:
Absolutní přírůstky (první diference řady) mají tvar

$$\Delta_t = y_t - y_{t-1}, \qquad t = 2,3,\dots,n ,$$

a tedy lze pro $t = 2, 3, \dots$ psát

$$\Delta_2 = y_2 - y_1 = 31586 - 29704 = 1882,$$
$$\Delta_3 = y_3 - y_2 = 31758 - 31586 = 172.$$
$$\dots$$

Relativní přírůstky jsou rovny

$$\delta_t = \frac{y_t - y_{t-1}}{y_{t-1}}, \qquad t = 2,3,\dots,n ,$$

a tedy lze pro $t = 2, 3, \dots$ psát

$$\delta_2 = \frac{y_2 - y_1}{y_1} = \frac{31586 - 29704}{29704} = 0{,}063,$$
$$\delta_3 = \frac{y_3 - y_2}{y_2} = \frac{31758 - 31586}{31586} = 0{,}005,$$
$$\dots$$

Koeficienty růstu jsou rovny

$$k_t = \frac{y_t}{y_{t-1}}, \qquad t = 2,3,\dots,n\,.$$

Pro t = 2, 3, … platí

$$k_2 = \frac{y_2}{y_1} = \frac{31586}{29704} = 1{,}063\,,$$

$$k_3 = \frac{y_3}{y_2} = \frac{31758}{31586} = 1{,}005\,,$$

$$\dots$$

V následující tabulce jsou uvedeny všechny hodnoty výše uvedených charakteristik.

rok	počty pracovníků	absolutní přírůstek	relativní přírůstek	koeficient růstu
2000	29 704	-	-	-
2001	31 586	1 882	0,063	1,063
2002	31 758	172	0,005	1,005
2003	32 824	1 066	0,034	1,034
2004	33 060	236	0,007	1,007
2005	31 288	-1 772	-0,054	0,946
2006	31 415	127	0,004	1,004
2007	31 129	-286	-0,009	0,991
2008	31 300	171	0,005	1,005
2009	29 133	-2 167	-0,069	0,931
2010	30 783	1 650	0,057	1,057
2011	28 113	-2 670	-0,087	0,913

Pro průměrný absolutní přírůstek platí

$$\overline{\Delta} = \frac{y_n - y_1}{n-1} = \frac{28113 - 29704}{11} = -144{,}636\,.$$

Průměrný koeficient růstu časové řady je roven

$$\overline{k} = \sqrt[n-1]{\frac{y_n}{y_1}} = \sqrt[11]{\frac{28113}{29704}} = 0{,}995\,.$$

Počet rozvodů tedy ročně poklesl v průměru o 0,5 %, což absolutně vyjádřeno, znamená průměrný roční pokles o 145 rozvodů.

Excel

V Excelu zadáme příslušné vzorce do jednotlivých buněk – jak výsledky tak vzorce jsou vidět na částečném výřezu tabulky.

	A	B	C	D	E
1	2000	29704			
2	2001	31586	1882	0,063	1,063
3	2002	31758	172	0,005	1,005
4	2003	32824	1066	0,034	1,034
5	2004	33060	236	0,007	1,007

	A	B	C	D	E
1	2000	29704			
2	2001	31586	=B2-B1	=(B2-B1)/B1	=B2/B1
3	2002	31758	=B3-B2	=(B3-B2)/B2	=B3/B2
4	2003	32824	=B4-B3	=(B4-B3)/B3	=B4/B3
5	2004	33060	=B5-B4	=(B5-B4)/B4	=B5/B4

7.2 Trendová analýza

7.2.1 Trendové křivky

Příklad 7.2
Roční časová řada v tabulce obsahuje údaje o průměrné výši mezd v České republice. Jedná se přitom o data za 2. čtvrtletí běžného roku, neboť v tomto čtvrtletí je nejstabilnější fond pracovní doby. Určete pro tuto řadu vhodnou trendovou křivku a odhadněte, jakou hodnotu bude mít průměrná mzda v roce 2012.

rok	t	průměrná mzda	rok	t	průměrná mzda
1995	1	8 311	2004	10	17 759
1996	2	9 962	2005	11	18 640
1997	3	11 322	2006	12	19 526
1998	4	12 026	2007	13	20 953
1999	5	12 982	2008	14	22 338
2000	6	13 541	2009	15	23 418

rok	t	průměrná mzda	rok	t	průměrná mzda
2001	7	14 743	2010	16	24 077
2002	8	15 964	2011	17	24 484
2003	9	17 748	2012	18	-

Zdroj: MPSV

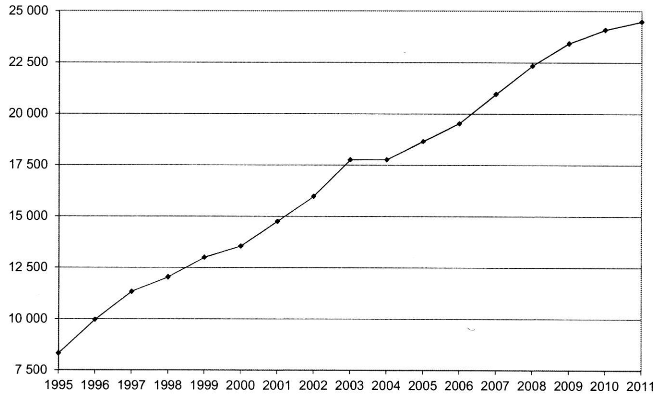
Řešení:
Nejprve se podíváme na grafický záznam řady. Z obrázku lze usoudit, že řada roste v podstatě lineárně a tak jako vhodnou trendovou křivku zvolíme přímku

$$T_t = \beta_0 + \beta_1 t, \qquad t = 1, 2, \ldots, n.$$

Odhadneme parametry přímky metodou nejmenších čtverců

$$b_1 = \frac{n\sum t\,y_t - \sum t \sum y_t}{n\sum t^2 - \left(\sum t\right)^2} = \frac{17 \cdot 3001899 - 153 \cdot 287\,794}{17 \cdot 1785 - 153^2} = 1009{,}2\,,$$

$$b_0 = \frac{\sum y_t}{n} - b_1 \frac{\sum t}{n} = \frac{287\,793}{17} - 1009{,}2 \cdot \frac{153}{17} = 7846{,}3\,.$$



Výsledná rovnice trendové funkce má tedy tvar

$$\hat{T}_t = 7846{,}3 + 1009{,}2 \cdot t\,.$$

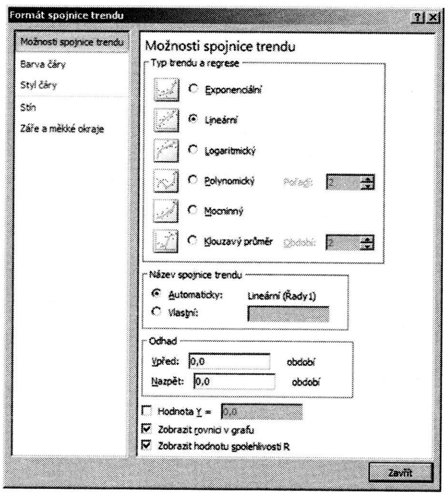
Odhad pro $t = 18$ (rok 2012) získáme dosazením do rovnice trendu, čímž obdržíme

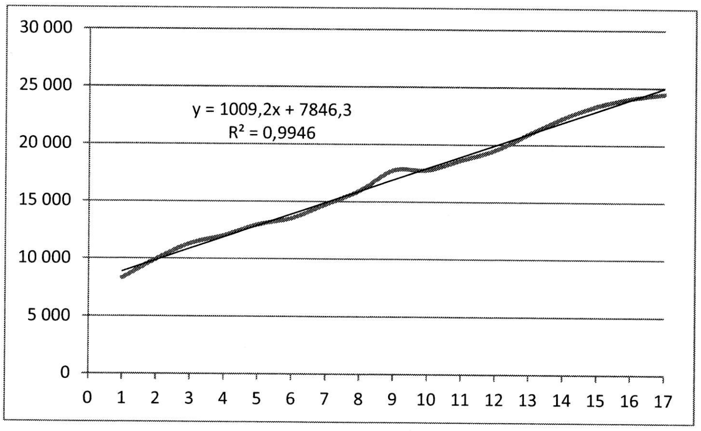
$$Y_{18} = \hat{T}_{18} = 7846{,}3 + 1009{,}2 \cdot 18 = 26011{,}9\,.$$

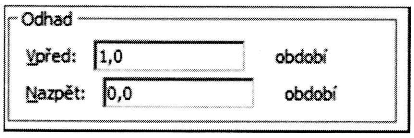
Pokud se tedy neočekávaně nezmění podmínky působící na vznik řady, naroste průměrná mzda za druhé čtvrtletí roku 2012 na hodnotu 26 011,90 Kč. Na druhé straně se jedná pouze o analytickou předpověď, která nebere v potaz ekonomické aspekty zkonstruované předpovědi.

Excel

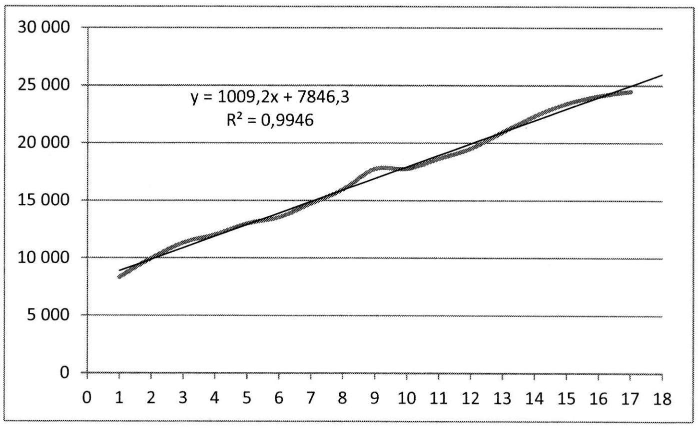
Do grafu časové řady můžeme v Excelu přidat spojnici trendu. Tím se nám zobrazí (pokud to v nabídce zaškrtneme) přímo v grafu rovnice regresní přímky. Je vidět, že výsledky této grafické analýzy odpovídají našim výpočtům.



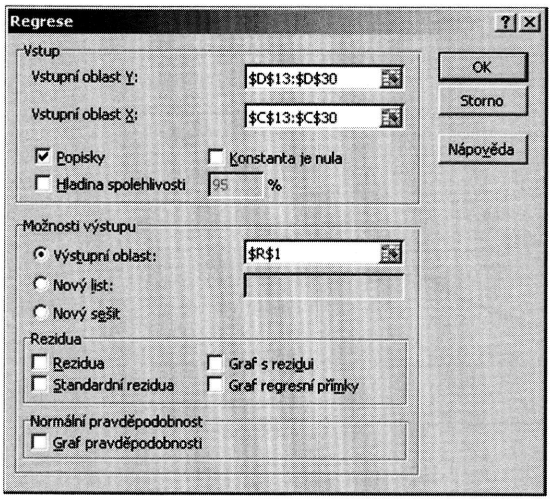




Odhad pro $t=18$ (rok 2012) můžeme získat i graficky tak, že v dialogovém okně *Formát spojnice trendu* v části *Odhad, Vpřed* vyplníme 1, čímž získáme předpověď - odhad trendové křivky o 1 období dopředu. Výsledek pak získáme pouze graficky.



Další možností je spočítat parametry trendové přímky pomocí nabídky Analýza dat a procedury Regrese. Do pole *Vstupní oblast Y* zadáme odkaz na hodnoty časové řady, do pole *Vstupní oblasti X* zadáme čas (vysvětlující proměnná) a zaškrtneme pole *Popisky* (nadpisy sloupců dat v Excelu).



	B	C	D
13	rok	t	Yt
14	1995	1	8 311
15	1996	2	9 962
16	1997	3	11 322
17	1998	4	12 026
18	1999	5	12 982
19	2000	6	13 541
20	2001	7	14 743
21	2002	8	15 964
22	2003	9	17 748
23	2004	10	17 759
24	2005	11	18 640
25	2006	12	19 526
26	2007	13	20 953
27	2008	14	22 338
28	2009	15	23 418
29	2010	16	24 077
30	2011	17	24 484

Obdržíme výstup z regresní analýzy, jehož součástí jsou i odhady parametrů b_0 (*Hranice*) a b_1 (t) jakož i index determinace (*Hodnota spolehlivosti R*) a korelační koeficient (*Násobné R*).

VÝSLEDEK

Regresní statistika	
Násobné R	0,99727
Hodnota spolehlivosti R	0,99456
Nastavená hodnota spolehlivosti R	0,99419
Chyba stř. hodnoty	389,36
Pozorování	17

ANOVA

	Rozdíl	*SS*	*MS*	*F*	*Významnost F*
Regrese	1	4,2E+08	4,2E+08	2741,01	2,1E-18
Rezidua	15	2274021	151601		
Celkem	16	4,2E+08			

	Koeficienty	*a stř. hodr*	*t Stat*	*Hodnota P*	*Dolní 95%*	*Horní 95%*	*Dolní 95,0%*	*Horní 95,0%*
Hranice	7846,27	197,522	39,7235	1,3E-16	7425,26	8267,28	7425,26	8267,28
t	1009,2	19,2762	52,3546	2,1E-18	968,112	1050,28	968,112	1050,28

□

Příklad 7.3

V tabulce jsou uvedeny hodnoty roční časové řady počtu narozených v České republice za období let 2006 - 2011. Vyrovnejte časovou řadu vhodnou trendovou funkcí a posuďte kvalitu vyrovnání.

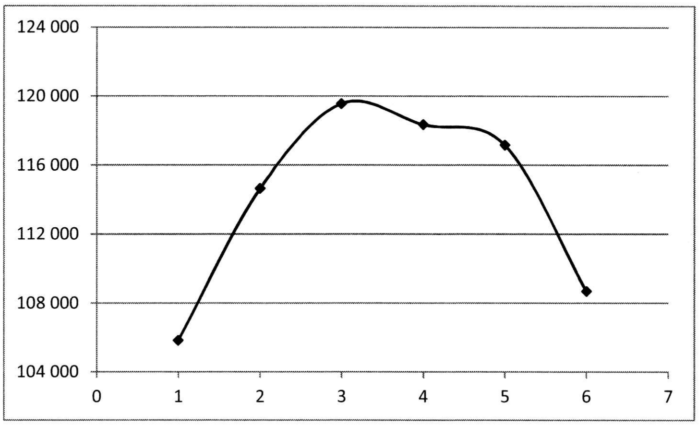
rok	t	počet narozených
2006	1	105 831
2007	2	114 632
2008	3	119 570
2009	4	118 348
2010	5	117 153
2011	6	108 673

Zdroj: http://www.czso.cz/

Řešení:

Nejprve se podíváme na grafický záznam řady. I když máme k dispozici poměrně málo pozorování, je přesto z obrázku jako vhodná trendová křivka patrná parabola, tedy polynom druhého stupně. Ten má tvar

$$T_t = \beta_0 + \beta_1 t + \beta_2 t^2, \qquad t = 1, 2, \ldots, n.$$



Odhadneme parametry paraboly metodou nejmenších čtverců, která vede k soustavě normálních rovnic ve tvaru

$$\sum y_t = n b_0 + b_1 \sum t + b_2 \sum t^2,$$
$$\sum t\, y_t = b_0 \sum t + b_1 \sum t^2 + b_2 \sum t^3,$$
$$\sum t^2 y_t = b_0 \sum t^2 + b_1 \sum t^3 + b_2 \sum t^4.$$

Po dosazení obdržíme soustavu rovnic ve tvaru

$$684\,207 = 6 \cdot b_0 + b_1 \cdot 21 + b_2 \cdot 91,$$
$$2\,405\,000 = b_0 \cdot 21 + b_1 \cdot 91 + b_2 \cdot 441,$$
$$10\,375\,110 = b_0 \cdot 91 + b_1 \cdot 441 + b_2 \cdot 2275.$$

Řešením této soustavy jsou odhady parametrů paraboly

$$b_0 = 93490, \qquad b_1 = 14454, \qquad b_2 = -1981.$$

Obdržíme tedy rovnici paraboly

$$\hat{T}_t = 93490 + 14454t - 1981t^2.$$

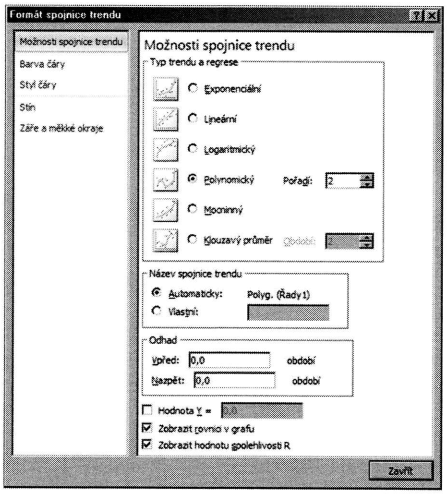
Vhodnost trendové křivky posoudíme pomocí indexu determinace (při použití metody nejmenších čtverců platí $\overline{Y} = \overline{y}$, což se projeví při zápisu v čitateli zlomku)

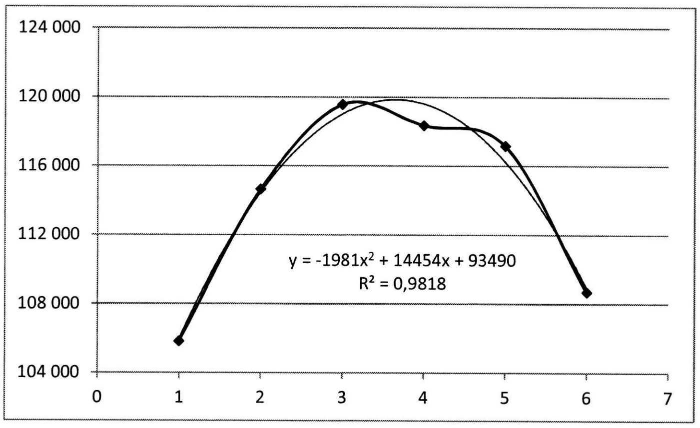
$$R^2 = \frac{S_T}{S_y} = \frac{\sum(\hat{Y}_t - \bar{y})^2}{\sum(y_t - \bar{y})^2} = \frac{152\ 545\ 599}{155\ 373\ 186} = 0{,}982.$$

Jako kritérium pro posouzení síly závislosti zvolíme (v časových řadách možná poněkud netradičně) index determinace. Ten je blízký 1, tudíž těsnost závislosti je velmi vysoká a parabola se tedy jeví jako vyhovující.

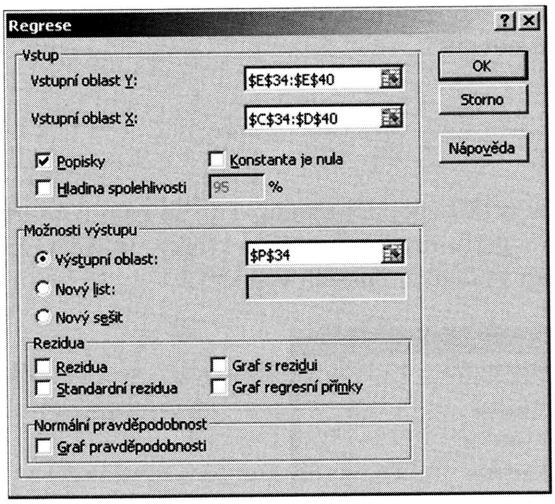
Excel

Do grafu časové řady můžeme v Excelu přidat spojnici trendu. Tím se nám zobrazí (pokud to v nabídce zaškrtneme) přímo v grafu rovnice trendové křivky. Je vidět, že výsledky této grafické analýzy odpovídají výsledkům našich výpočtů.





Další možností je spočítat parametry trendové přímky pomocí nabídky **Analýza dat** a procedury **Regrese**. Postupujeme obdobně jako v případě přímky. Jelikož však trendovou křivkou je nyní parabola, musíme proměnnou t^2 (tedy čtverec vysvětlující proměnné) nejprve v Excelu dopočítat.



	C	D	E
34	t	t^2	Y_t
35	1	1	105 831
36	2	4	114 632
37	3	9	119 570
38	4	16	118 348
39	5	25	117 153
40	6	36	108 673

Výstup z regresní analýzy opět odpovídá našim předchozím výpočtům.

VÝSLEDEK

Regresní statistika	
Násobné R	0,99086
Hodnota spolehlivosti R	0,9818
Nastavená hodnota spolehlivosti R	0,96967
Chyba stř. hodnoty	970,839
Pozorování	6

ANOVA

	Rozdíl	*SS*	*MS*	*F*	*Významnost F*
Regrese	2	152545598,7	7,6E+07	80,9236	0,00246
Rezidua	3	2827586,807	942529		
Celkem	5	155373185,5			

	Koeficienty	*Chyba stř. hodnoty*	*t Stat*	*Hodnota P*	*Dolní 95%*	*Horní 95%*	*Dolní 95,(*	*Horní 95,0%*
Hranice	93489,9	1736,690126	53,8322	1,4E-05	87963	99016,8	87963	99016,8
t	14454,3	1136,190136	12,7217	0,00105	10838,4	18070,2	10838,4	18070,2
t2	-1981,02	158,890877	-12,4678	0,00111	-2486,68	-1475,4	-2486,7	-1475,36

Příklad 7.4

Firma, zabývající se internetovým prodejem, se v posledních 10 letech velmi rychle rozvíjela. V tabulce jsou údaje o tržbách firmy v mil. Kč za roky 2003 – 2012. Najděte vhodnou trendovou křivku, kterou data vyrovnáte, a odhadněte vývoj tržeb na 2 roky dopředu.

rok	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
tržby	102	113	149	189	202	256	291	378	422	515

Zdroj: vlastní data

Řešení:
Nejprve se podíváme na koeficienty růstu této časové řady.

t	2	3	4	5	6	7	8	9	10
k_t	1,11	1,32	1,27	1,07	1,27	1,14	1,30	1,12	1,22

Z tabulky je patrné, že koeficienty růstu kolísají kolem určité konstantní úrovně, což poukazuje na exponenciálu jako na vhodnou trendovou křivku.

$$T_t = \beta_0 \beta_1^t, \quad t = 1, 2, \ldots, n.$$

Po zlogaritmování obdržíme

$$\ln T_t = \ln \beta_0 + t \ln \beta_1, \quad t = 1, 2, \ldots, n.$$

Nyní můžeme odhadnout parametry exponenciály metodou nejmenších čtverců

$$\ln b_1 = \frac{n\sum t \ln y_t - \sum t \sum \ln y_t}{n\sum t^2 - \left(\sum t\right)^2} = \frac{10 \cdot 313{,}91 - 55 \cdot 54{,}35}{10 \cdot 385 - 55^2} = 0{,}181636,$$

$$\ln b_0 = \frac{\sum \ln y_t}{n} - \ln b_1 \frac{\sum t}{n} = \frac{54{,}348889}{10} - 0{,}181636 \cdot \frac{55}{10} = 4{,}435891.$$

Pro výsledný model tedy můžeme psát

$$\ln \hat{T}_t = 4{,}435891 + 0{,}181636 \cdot t,$$

neboli

$$\hat{T}_t = \exp(4{,}435891 + 0{,}181636 \cdot t) = 84{,}427316 \cdot e^{0{,}181636 \cdot t} = 84{,}427316 \cdot 1{,}199176^t.$$

Předpovědi pro $t = 11$ (rok 2013) a $t = 12$ (rok 2014) dostaneme dosazením za t do rovnice trendu

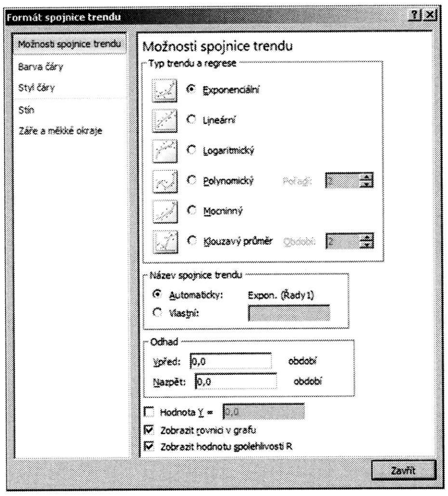
$$\hat{T}_{11} = 84,427316 \cdot 1,199176^{11} = 622,580,$$

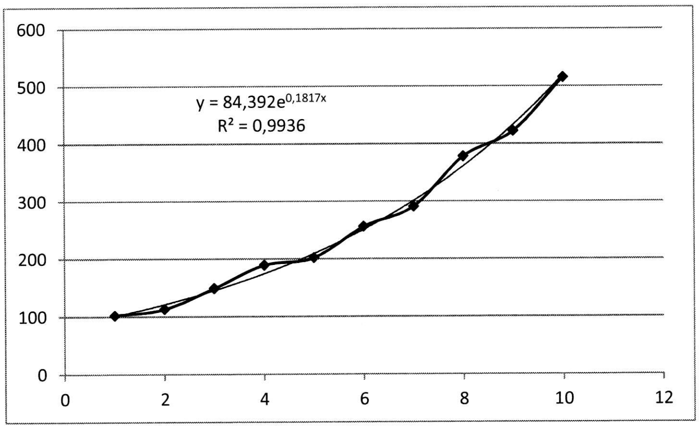
$$\hat{T}_{12} = 84,427316 \cdot 1,199176^{12} = 746,583.$$

Firma tedy může očekávat v roce 2013 tržby ve výši necelých 623 mil. Kč a v roce 2014 tržby ve výši necelých 747 mil. Kč.

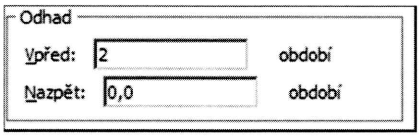
Excel

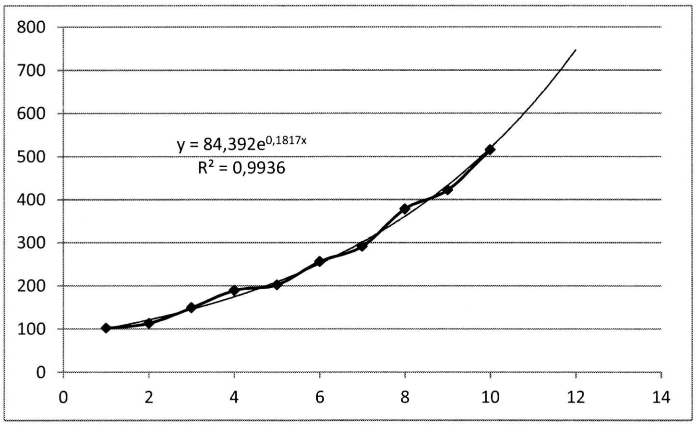
Do grafu časové řady můžeme v Excelu přidat spojnici trendu. Tím se nám zobrazí (pokud to v nabídce zaškrtneme) přímo v grafu rovnice trendové křivky. Je vidět, že výsledky této grafické analýzy přesně odpovídají výsledkům našich výpočtů.





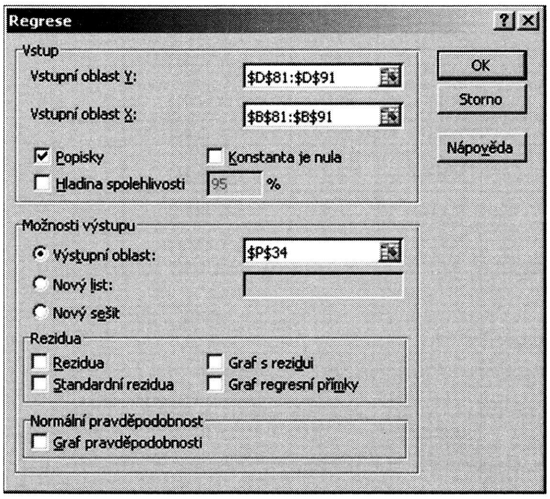
Dosažené výsledky této grafické analýzy jsou plně v souladu s výsledky našich výpočtů. Zadáme-li v dialogovém okně *Formát* spojnice trendu v části *Odhad*, *Vpřed* číslo 2, získáme odhad trendové křivky o 2 období dopředu. Výsledek pak vidíme znázorněny graficky.





Nabízí se opět další možnost výpočtu parametrů trendové přímky a to pomocí nabídky **Analýza dat** a procedury **Regrese**. Postupujeme obdobně jako v předchozím příkladu.

Trendovou křivkou je nyní exponenciála, a musíme tudíž dopočítat proměnnou $\ln Y_t$ (tedy logaritmus vysvětlované proměnné), kterou dosadíme jako vysvětlovanou proměnnou. Jako trendovou křivku použijeme přímku s tím, že odhadnuté hodnoty jejích parametrů je dále třeba odlogaritmovat



	B	C	D
81	t	Y_t	$\ln Y_t$
82	1	102	4,6249728
83	2	113	4,7273878
84	3	149	5,0039463
85	4	189	5,241747
86	5	202	5,3082677
87	6	256	5,5451774
88	7	291	5,6733233
89	8	378	5,9348942
90	9	422	6,0450053
91	10	515	6,2441669

VÝSLEDEK

Regresní statistika	
Násobné R	0,99677
Hodnota spolehlivosti R	0,99355
Nastavená hodnota spolehlivosti R	0,99275
Chyba stř. hodnoty	0,04701
Pozorování	10

ANOVA

	Rozdíl	SS	MS	F	íznamnost F
Regrese	1	2,724082244	2,72408	1232,58	4,7E-10
Rezidua	8	0,017680567	0,00221		
Celkem	9	2,741762811			

	Koeficienty	Chyba stř. hodnoty	t Stat	Hodnota P	Dolní 95%	Horní 95%	olní 95,0%	Horní 95,0%
Hranice	4,43547	0,032114894	138,113	8,4E-15	4,36142	4,50953	4,36142	4,50953
t	0,18171	0,005175784	35,1081	4,7E-10	0,16978	0,19365	0,16978	0,19365

Tím jsme obdrželi model $\ln \hat{T}_t = 4,43547 + 0,18171 \cdot t$, který zcela odpovídá již dosaženým výsledkům. Dále je postup již zřejmý.

◘

7.2.2 Klouzavé průměry

Příklad 7.5
V tabulce jsou hodnoty časové řady počtu zahájených bytů v rodinných domech v ČR za období let 1998 - 2011. Vyrovnejte tuto řadu jednoduchými klouzavými průměry délky 3, 5 a 7.

rok	t	počet bytů	rok	t	počet bytů
1998	1	14 933	2005	8	17 579
1999	2	12 489	2006	9	20 620
2000	3	12 177	2007	10	20 990
2001	4	12 895	2008	11	22 918
2002	5	13 659	2009	12	18 750
2003	6	17 250	2010	13	16 611
2004	7	17 485	2011	14	17 060

Zdroj: http://www.czso.cz/

Řešení:
3-členný klouzavý průměr pro rok 1999 je

$$\overline{y}_2 = \frac{y_1 + y_2 + y_3}{3} = \frac{14933 + 12489 + 12177}{3} = 13200,$$

pro rok 2000

$$\overline{y}_3 = \frac{y_2 + y_3 + y_4}{3} = \frac{12\,489 + 12\,177 + 12\,895}{3} = 12\,520\,,$$

atd.

5-členný klouzavý průměr pro rok 2000 je

$$\overline{y}_3 = \frac{y_1 + y_2 + y_3 + y_4 + y_5}{5} = \frac{14\,933 + 12\,489 + 12\,177 + 12\,895 + 13\,659}{5} = 13\,231,$$

pro 2001

$$\overline{y}_4 = \frac{y_2 + y_3 + y_4 + y_5 + y_6}{5} = \frac{12\,489 + 12\,177 + 12\,895 + 13\,659 + 17\,250}{5} = 13\,694,$$

atd.

7-členný klouzavý průměr pro rok 2001 je

$$\overline{y}_4 = \frac{y_1 + y_2 + y_3 + y_4 + y_5 + y_6 + y_7}{7} =$$

$$= \frac{14\,933 + 12\,489 + 12\,177 + 12\,895 + 13\,659 + 17\,250 + 17\,485}{7} = 14\,413\,,$$

pro rok 2002

$$\overline{y}_5 = \frac{y_2 + y_3 + y_4 + y_5 + y_6 + y_7 + y_8}{7} =$$

$$= \frac{12\,489 + 12\,177 + 12\,895 + 13\,659 + 17\,250 + 17\,485 + 17\,579}{7} = 14\,791,$$

atd.

V následující tabulce uvádíme všechny hodnoty klouzavých průměrů.

rok	počet bytů	klouzavé průměry délky 3	klouzavé průměry délky 5	klouzavé průměry délky 7
1.	14 933	-	-	-
2.	12 489	13 200	-	-

rok	počet bytů	klouzavé průměry délky 3	klouzavé průměry délky 5	klouzavé průměry délky 7
3.	12 177	12 520	13 231	-
4.	12 895	12 910	13 694	14 413
5.	13 659	14 601	14 693	14 791
6.	17 250	16 131	15 774	15 952
7.	17 485	17 438	17 319	17 211
8.	17 579	18 561	18 785	18 643
9.	20 620	19 730	19 918	19 370
10.	20 990	21 509	20 171	19 279
11.	22 918	20 886	19 978	19 218
12.	18 750	19 426	19 266	-
13.	16 611	17 474	-	-
14.	17 060	-	-	-

Excel

V Excelu si snadno poradíme zadáním příslušných vzorců – vše je vidět na částečném výřezu tabulky. Pro ilustraci jsou použity jak vzorce, tak funkce *Suma*.

	A	B	C	D	E
1	1998	14 933			
2	1999	12 489	13 200		
3	2000	12 177	12 520	13 231	
4	2001	12 895	12 910	13 694	14 413
5	2002	13 659	14 601	14 693	14 791
6	2003	17 250	16 131	15 774	15 952

	A	B	C	D	E
1	1998	14933			
2	1999	12489	=(B1+B2+B3)/3		
3	2000	12177	=(B2+B3+B4)/3	=SUMA(B1:B5)/5	
4	2001	12895	=(B3+B4+B5)/3	=SUMA(B2:B6)/5	=SUMA(B1:B7)/7
5	2002	13659	=(B4+B5+B6)/3	=SUMA(B3:B7)/5	=SUMA(B2:B8)/7
6	2003	17250	=(B5+B6+B7)/3	=SUMA(B4:B8)/5	=SUMA(B3:B9)/7

Příklad 7.6

V tabulce jsou uvedeny hodnoty časové řady výroby piva (v tis. hl) v České republice za roky 1993 - 2009. Vyrovnejte tuto časovou řadu

a) 5-člennými klouzavými průměry 2. řádu,

b) 7-člennými klouzavými průměry 2. řádu.

rok	t	pivo (tis. hl)	rok	t	pivo (tis. hl)
1993	1	17 366	2002	10	17 987
1994	2	17 876	2003	11	18 216
1995	3	17 687	2004	12	18 596
1996	4	18 057	2005	13	18 885
1997	5	18 558	2006	14	20 134
1998	6	18 290	2007	15	18 627
1999	7	17 946	2008	16	19 213
2000	8	17 796	2009	17	17 813
2001	9	17 734			

Zdroj: http://www.czso.cz/

Řešení:

a) 5-členný klouzavý průměr 2. řádu má váhy $\frac{1}{35}(-3,12,17,12,-3)$.

Potom první vyrovnaná hodnota pro $t = 3$ je rovna

$$\overline{y}_3 = \frac{-3y_1 + 12y_2 + 17y_3 + 12y_4 - 3y_5}{35} =$$

$$= \frac{-3\cdot 17366 + 12\cdot 17876 + 17\cdot 17687 + 12\cdot 18057 - 3\cdot 18558}{35} = 17\,831{,}5\,.$$

Druhá vyrovnaná hodnota pro $t = 4$ je rovna

$$\overline{y}_4 = \frac{-3y_2 + 12y_3 + 17y_4 + 12y_5 - 3y_6}{35} =$$

$$= \frac{-3\cdot 17876 + 12\cdot 17687 + 17\cdot 18057 + 12\cdot 18558 - 3\cdot 18290}{35} = 18\,097{,}5\,,$$

atd.

b) 7-členný klouzavý průměr 2. řádu má váhy $\frac{1}{21}(-2,3,6,7,6,3,-2)$. Potom první vyrovnaná hodnota pro $t = 4$ je rovna

$$\overline{y}_4 = \frac{-2y_1 + 3y_2 + 6y_3 + 7y_4 + 6y_5 + 3y_6 - 2y_7}{21} =$$

$$= \frac{-2 \cdot 17366 + 3 \cdot 17876 + 6 \cdot 17687 + 7 \cdot 18057 + 6 \cdot 18558 + 3 \cdot 18290 - 2 \cdot 17946}{21} =$$

$$= 18\ 178{,}2.$$

Vyrovnaná hodnota pro $t = 5$ je rovna

$$\overline{y}_5 = \frac{-2y_2 + 3y_3 + 6y_4 + 7y_5 + 6y_6 + 3y_7 - 2y_8}{21} =$$

$$= \frac{-2 \cdot 17876 + 3 \cdot 17687 + 6 \cdot 18057 + 7 \cdot 18558 + 6 \cdot 18290 + 3 \cdot 17946 - 2 \cdot 17796}{21} =$$

$$= 18\ 264{,}0\ ,$$

atd.

rok	Y_t	5-členné klouzavé průměry 2. řádu	7-členné klouzavé průměry 2. řádu
1993	17 366	-	-
1994	17 876	-	-
1995	17 687	17 831,5	-
1996	18 057	18 097,5	18 178,2
1997	18 558	18 421,5	18 264,0
1998	18 290	18 326,3	18 274,8
1999	17 946	17 978,2	18 044,1
2000	17 796	17 767,5	17 806,4
2001	17 734	17 782,5	17 788,1
2002	17 987	17 942,9	17 958,2
2003	18 216	18 251,7	18 143,2
2004	18 596	18 485,2	18 781,9
2005	18 885	19 293,6	19 081,1
2006	20 134	19 399,9	19 399,0
2007	18 627	19 392,3	-
2008	19 213	-	-
2009	17 813	-	-

V předchozí tabulce jsou uvedeny hodnoty všech 5-členných i 7-členných klouzavých průměrů řádu 2.

Excel

V Excelu si snadno získáme požadované výsledky zadáním příslušných vzorců – vše je vidět na částečném výřezu tabulky.

	A	B	C	D	E
1	rok	t	pivo	5členný	7členný
2			(tis. hl)		
3	1993	1	17 366	-	-
4	1994	2	17 876	-	-
5	1995	3	17 687	17 831,5	-
6	1996	4	18 057	18 097,5	18 178,2
7	1997	5	18 558	18 421,5	18 264,0
8	1998	6	18 290	18 326,3	18 274,8

	A	B	C	D	E
1	rok	t	pivo	5členný	7členný
2			(tis. hl)		
3	1993	1	17366	-	-
4	1994	2	17876	-	-
5	1995	3	17687	=(-3*C3+12*C4+17*C5+12*C6-3*C7)/35	-
6	1996	4	18057	=(-3*C4+12*C5+17*C6+12*C7-3*C8)/35	=(-2*C3+3*C4+6*C5+7*C6+6*C7+3*C8-2*C9)/21
7	1997	5	18558	=(-3*C5+12*C6+17*C7+12*C8-3*C9)/35	=(-2*C4+3*C5+6*C6+7*C7+6*C8+3*C9-2*C10)/21
8	1998	6	18290	=(-3*C6+12*C7+17*C8+12*C9-3*C10)/35	=(-2*C5+3*C6+6*C7+7*C8+6*C9+3*C10-2*C11)/21
9	1999	7	17946	=(-3*C7+12*C8+17*C9+12*C10-3*C11)/35	=(-2*C6+3*C7+6*C8+7*C9+6*C10+3*C11-2*C12)/21
10	2000	8	17796	=(-3*C8+12*C9+17*C10+12*C11-3*C12)/35	=(-2*C7+3*C8+6*C9+7*C10+6*C11+3*C12-2*C13)/21

▫

7.2.3 Exponenciální vyrovnávání

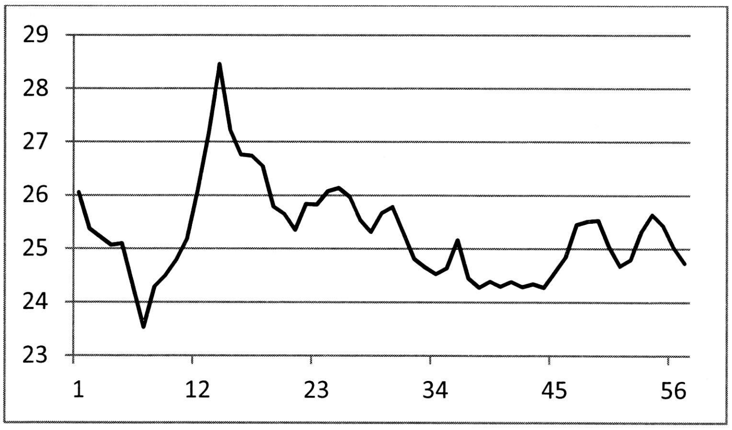
Příklad 7.7

V následující tabulce jsou hodnoty měsíčních průměrných kurzů české koruny vůči Euru za období leden 2008 - září 2012 (proměnná *EUR*). Pomocí jednoduchého exponenciálního vyrovnávání se zvolenou vyrovnávací konstantou $\alpha = 0{,}9$ odhadněte hodnotu této časové řady pro říjen roku 2012 a poté na Internetu na webových stránkách ČNB (www.cnb.cz) ověřte kvalitu předpovědi (vysoká hodnota $\alpha = 0{,}9$ znamená, že při vyrovnávání časové řady přisuzujeme poměrně velký význam poslední známé hodnotě).

2008	*EUR*	2009	*EUR*	2010	*EUR*	2011	*EUR*	2012	*EUR*
1	26,051	1	27,169	1	26,136	1	24,449	1	25,532
2	25,376	2	28,459	2	25,976	2	24,276	2	25,041
3	25,221	3	27,229	3	25,540	3	24,392	3	24,676
4	25,067	4	26,760	4	25,313	4	24,291	4	24,799
5	25,098	5	26,738	5	25,666	5	24,383	5	25,322
6	24,314	6	26,545	6	25,780	6	24,285	6	25,641
7	23,529	7	25,787	7	25,305	7	24,341	7	25,434
8	24,286	8	25,649	8	24,807	8	24,273	8	25,020
9	24,497	9	25,349	9	24,651	9	24,557	9	24,731
10	24,787	10	25,836	10	24,526	10	24,848		
11	25,183	11	25,827	11	24,637	11	25,453		
12	26,106	12	26,076	12	25,165	12	25,515		

Zdroj: http://www.cnb.cz/

Podívejme se nejprve na průběh této analyzované řady na obrázku. Je zřejmé, že pro tuto řadu nelze nalézt vhodnou trendovou křivku v celé její délce, je tedy namístě pokusit se modelovat trend adaptivně pomocí exponenciálního vyrovnávání.



Při výpočtech budeme používat obvyklý rekurentní vztah

$$Y_t = \alpha\, y_t + (1-\alpha)Y_{t-1}\,.$$

Podotkněme ještě, že někdy se předchozí vztah vyjadřuje ve tvaru

$$Y_t = (1-\alpha)\,y_t + \alpha Y_{t-1}\,.$$

což je v podstatě totéž (stačí zaměnit α za $1-\alpha$). Chybu předpovědi počítáme dle vzorce

$$E_t = y_t - Y_t(t-1)\,,$$

kde $Y_t(t-1)$ je předpověď hodnoty y_t, konstruovaná v čase $t-1$.

Potřebujeme získat tzv. „startovací" hodnotu y_0, přičemž je zřejmé, že v bodu $t=0$ žádné pozorování nemáme k dispozici. Tato situace se dá řešit nejrůznějším způsobem – například lze jako startovací hodnotu zvolit průměr z několika prvních pozorování, ale existují i další možnosti, jak dále uvidíme. V našem příkladu zvolíme za startovací hodnotu průměr z prvních osmi pozorování – tedy hodnotu $y_0 = 24{,}868$.
Jako kritérium pro výběr vhodné vyrovnávací konstanty nám bude sloužit statistika SSE (součet střední čtvercové chyby) ve tvaru

$$\mathrm{SSE} = \sum (y_t - Y_t(t-1))^2\,.$$

Pak už pro $\alpha = 0{,}9$ snadno vypočítáme

$$Y_1 = 0{,}9 \cdot 26{,}051 + (1-0{,}9)\cdot 24{,}868 = 25{,}933\,,$$
$$Y_2 = 0{,}9 \cdot 25{,}376 + (1-0{,}9)\cdot 25{,}933 = 25{,}432\,,$$
$$\ldots$$

Pro $\alpha = 0{,}9$ je statistika SSE rovna

$$\mathrm{SSE} = \sum (y_t - Y_t(t-1))^2 = 14{,}407\,.$$

V následující tabulce jsou uvedeny všechny vyrovnané hodnoty a z nich plynoucí výpočty.

Rok	Měsíc	y_t	Y_t	$Y_t(t-1)$	$y_t - Y_t(t-1)$	$(y_t - Y_t(t-1))^2$
2008	Leden	26,051	25,933	24,868	1,183	1,400
	Únor	25,376	25,432	25,933	-0,557	0,310
	Březen	25,221	25,242	25,432	-0,211	0,044

Rok	Měsíc	y_t	Y_t	$Y_t(t-1)$	$y_t - Y_t(t-1)$	$(y_t - Y_t(t-1))^2$
	Duben	25,067	25,085	25,242	-0,175	0,031
	Květen	25,098	25,097	25,085	0,013	0,000
	Červen	24,314	24,392	25,097	-0,783	0,613
	Červenec	23,529	23,615	24,392	-0,863	0,745
	Srpen	24,286	24,219	23,615	0,671	0,450
	Září	24,497	24,469	24,219	0,278	0,077
	Říjen	24,787	24,755	24,469	0,318	0,101
	Listopad	25,183	25,140	24,755	0,428	0,183
	Prosinec	26,106	26,009	25,140	0,966	0,933
2009	Leden	27,169	27,053	26,009	1,160	1,345
	Únor	28,459	28,318	27,053	1,406	1,977
	Březen	27,229	27,338	28,318	-1,089	1,187
	Duben	26,760	26,818	27,338	-0,578	0,334
	Květen	26,738	26,746	26,818	-0,080	0,006
	Červen	26,545	26,565	26,746	-0,201	0,040
	Červenec	25,787	25,865	26,565	-0,778	0,605
	Srpen	25,649	25,671	25,865	-0,216	0,047
	Září	25,349	25,381	25,671	-0,322	0,103
	Říjen	25,836	25,791	25,381	0,455	0,207
	Listopad	25,827	25,823	25,791	0,036	0,001
	Prosinec	26,076	26,051	25,823	0,253	0,064
2010	Leden	26,136	26,127	26,051	0,085	0,007
	Únor	25,976	25,991	26,127	-0,151	0,023
	Březen	25,540	25,585	25,991	-0,451	0,204
	Duben	25,313	25,340	25,585	-0,272	0,074
	Květen	25,666	25,633	25,340	0,326	0,106
	Červen	25,780	25,765	25,633	0,147	0,021
	Červenec	25,305	25,351	25,765	-0,460	0,212

Rok	Měsíc	y_t	Y_t	$Y_t(t-1)$	$y_t - Y_t(t-1)$	$(y_t - Y_t(t-1))^2$
	Srpen	24,807	24,861	25,351	-0,544	0,296
	Září	24,651	24,672	24,861	-0,210	0,044
	Říjen	24,526	24,541	24,672	-0,146	0,021
	Listopad	24,637	24,627	24,541	0,096	0,009
	Prosinec	25,165	25,111	24,627	0,538	0,289
2011	Leden	24,449	24,515	25,111	-0,662	0,439
	Únor	24,276	24,300	24,515	-0,239	0,057
	Březen	24,392	24,383	24,300	0,092	0,008
	Duben	24,291	24,300	24,383	-0,092	0,008
	Květen	24,383	24,375	24,300	0,083	0,007
	Červen	24,285	24,294	24,375	-0,090	0,008
	Červenec	24,341	24,336	24,294	0,047	0,002
	Srpen	24,273	24,279	24,336	-0,063	0,004
	Září	24,557	24,529	24,279	0,278	0,077
	Říjen	24,848	24,816	24,529	0,319	0,102
	Listopad	25,453	25,389	24,816	0,637	0,406
	Prosinec	25,515	25,502	25,389	0,126	0,016
2012	Leden	25,532	25,529	25,502	0,030	0,001
	Únor	25,041	25,090	25,529	-0,488	0,238
	Březen	24,676	24,717	25,090	-0,414	0,171
	Duben	24,799	24,791	24,717	0,082	0,007
	Květen	25,322	25,269	24,791	0,531	0,282
	Červen	25,641	25,604	25,269	0,372	0,138
	Červenec	25,434	25,451	25,604	-0,170	0,029
	Srpen	25,020	25,063	25,451	-0,431	0,186
	Září	24,731	24,764	25,063	-0,332	0,110
	celkem					14,407

Protože u jednoduchého exponenciálního vyrovnávání je nejlepší předpovědí poslední vyrovnaná hodnota, můžeme odhad kurzu Kč/EUR pro říjen 2012 položit roven hodnotě 25,063. Ve skutečnosti byl říjnový průměr kurzů roven číslu 24,955.

Excel

Zadání včetně použitých vzorců je znázorněno na následujících dvou obrázcích:

	E	F	G	H	I	K	M	N	O
1	**průměr**	t			y_t	Y_t	$Y_t(t-1)$	$y_t - Y_t(t-1)$	$(y_t - Y_t(t-1))^2$
2	24,8678	1	2008	Leden	26,051	25,933	24,868	1,183	1,400
3		2		Únor	25,376	25,432	25,933	-0,557	0,310
4	α	3		Březen	25,221	25,242	25,432	-0,211	0,044
5	0,9	4		Duben	25,067	25,085	25,242	-0,175	0,031
6		5		Květen	25,098	25,097	25,085	0,013	0,000

	E	F	G	H	I	K	M	N	O
1	**průměr**	t			y_t	Y_t	$Y_t(t-1)$	$y_t - Y_t(t-1)$	$(y_t - Y_t(t-1))^2$
2	=PRŮMĚR(I2:I9)	1	2008	Leden	26,051	=$E$5*I2+(1-$E$5)*E2	=E2	=I2-M2	=N2*N2
3		2		Únor	25,376	=$E$5*I3+(1-$E$5)*K2	=M2+$E$5*(I2-M2)	=I3-M3	=N3*N3
4	α	3		Březen	25,221	=$E$5*I4+(1-$E$5)*K3	=M3+$E$5*(I3-M3)	=I4-M4	=N4*N4
5	0,9	4		Duben	25,067	=$E$5*I5+(1-$E$5)*K4	=M4+$E$5*(I4-M4)	=I5-M5	=N5*N5
6		5		Květen	25,098	=$E$5*I6+(1-$E$5)*K5	=M5+$E$5*(I5-M5)	=I6-M6	=N6*N6

7.3 Sezónnost v časových řadách

7.3.1 Regresní přístup k sezónní složce

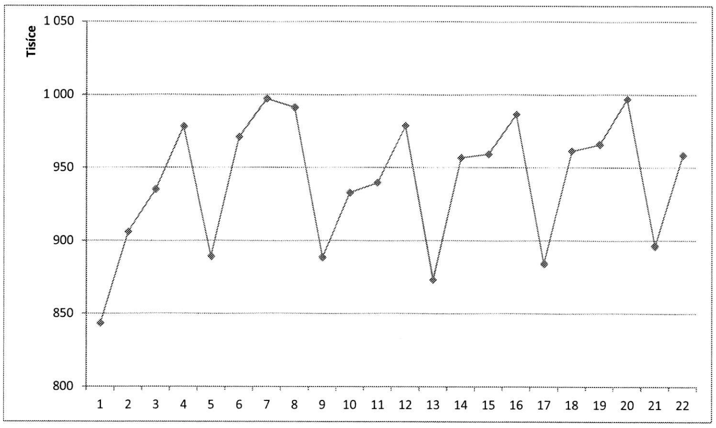
Příklad 7.8

Uvedená tabulka obsahuje hodnoty čtvrtletní časové řady výdajů na hrubý domácí produkt ČR v běžných cenách v mil. Kč v letech 2007 - 2012. Modelujte trendovou a sezónní složku této řady pomocí regresního přístupu a odhadněte vývoj této časové řady na poslední dvě čtvrtletí roku 2012 a na dvě první čtvrtletí roku 2013. Pokud jsou již známy hodnoty za tento rok, porovnejte vypočtené předpovědi se skutečností. Zdrojem dat jsou webové stránky Českého statistického úřadu, kde můžete kvalitu vypočítaných předpovědí ověřit.

rok/čtvrtletí	1	2	3	4
2007	843 399	905 917	935 100	978 157
2008	889 080	970 995	997 237	991 099
2009	888 452	932 655	939 543	978 575
2010	872 980	956 630	959 164	986 463
2011	884 085	961 173	965 752	996 792
2012	896 108	958 377		

Zdroj: http://www.czso.cz/

Řešení:
Nejprve se podíváme na grafický záznam řady na následujícím obrázku. Z něho je patrné, že sledovaná časová řada vykazuje mírně rostoucí lineární trend a aditivní sezónnost.



Jako vhodný model zvolíme

$$y_t = \beta_0 + \beta_1 t + \alpha_1 x_{1t} + \alpha_2 x_{2t} + \alpha_3 x_{3t} + \varepsilon_t ,$$

kde x_{1t}, x_{2t}, x_{3t} jsou nula-jedničkové proměnné detekující 1., 2. nebo 3. čtvrtletí, definované jako

$$\begin{aligned} x_{it} &= 1, \quad \text{pro } t \text{ odpovídající } i\text{-tému čtvrtletí v roce,} \\ &= 0, \quad \text{jinak,} \end{aligned}$$

pro $i = 1, 2, 3$ (pracujeme se čtvrtletní řadou). Konkrétní hodnoty x_{it} jsou uvedeny v následující tabulce.

Na základě odhadů parametrů α_1, α_2 a α_3 pak vypočítáme sezónní faktory. Do modelu není zařazen parametr α_4 (4. čtvrtletí). Ten však pro výpočet sezónního faktoru nepotřebujeme. Při jeho určení vycházíme z podmínky jednoznačnosti dekompozičního rozkladu, při které se požaduje, aby se vliv sezónních faktorů v rámci každého roku celkově vykompenzoval. V důsledku tohoto požadavku musí být součet sezónních aditivních faktorů roven 0. Tato podmínka již jednoznačně určuje hodnotu sezónního faktoru pro 4. čtvrtletí.

Pro odhad parametrů modelu aplikujeme metodu nejmenších čtverců a získáme odhady b_0, b_1, a_1, a_2, a_3 parametrů $\beta_0, \beta_1, \alpha_1, \alpha_2, \alpha_3$ ve tvaru

$$b_0 = 969809, \quad b_1 = 1367,$$
$$a_1 = -105833, \quad a_2 = -38593, \quad a_3 = -25491.$$

Protože součet aditivních sezónních faktorů musí být roven nule (uvažujeme zaokrouhlení), vypočítáme jednotlivé sezónní faktory následujícím způsobem:

$$\bar{a} = \frac{a_1 + a_2 + a_3}{4} = \frac{-105833 - 38593 - 25491}{4} = -42479,$$
$$S_{1+4j} = a_1 - \bar{a} = -105833 - (-42479) = -63354,$$
$$S_{2+4j} = a_2 - \bar{a} = -38593 - (-42479) = 3886,$$
$$S_{3+4j} = a_3 - \bar{a} = -25491 - (-42479) = 16988,$$
$$S_{4+4j} = -\bar{a} = 42479.$$

Potom přepočítáme parametry lineárního trendu

t	y_t	x_{1t}	x_{2t}	x_{3t}	x_{4t}	$Y_t = T_t + S_t$	S_t	T_t	$y_t - Y_t$
1	843 399	1	0	0	0	865 344	-63 354	928 698	-21 945
2	905 917	0	1	0	0	933 951	3 886	930 065	-28 034
3	935 100	0	0	1	0	948 421	16 988	931 432	-13 321
4	978 157	0	0	0	1	975 279	42 479	932 800	2 878
5	889 080	1	0	0	0	870 813	-63 354	934 167	18 267
6	970 995	0	1	0	0	939 420	3 886	935 534	31 575

t	y_t	x_{1t}	x_{2t}	x_{3t}	x_{4t}	$Y_t = T_t + S_t$	S_t	T_t	$y_t - Y_t$
7	997 237	0	0	1	0	953 890	16 988	936 902	43 347
8	991 099	0	0	0	1	980 748	42 479	938 269	10 351
9	888 452	1	0	0	0	876 283	-63 354	939 636	12 169
10	932 655	0	1	0	0	944 890	3 886	941 004	-12 235
11	939 543	0	0	1	0	959 359	16 988	942 371	-19 816
12	978 575	0	0	0	1	986 217	42 479	943 738	-7 642
13	872 980	1	0	0	0	881 752	-63 354	945 106	-8 772
14	956 630	0	1	0	0	950 359	3 886	946 473	6 271
15	959 164	0	0	1	0	964 829	16 988	947 840	-5 665
16	986 463	0	0	0	1	991 687	42 479	949 208	-5 224
17	884 085	1	0	0	0	887 221	-63 354	950 575	-3 136
18	961 173	0	1	0	0	955 829	3 886	951 942	5 344
19	965 752	0	0	1	0	970 298	16 988	953 310	-4 546
20	996 792	0	0	0	1	997 156	42 479	954 677	-364
21	896 108	1	0	0	0	892 691	-63 354	956 044	3 417
22	958 377	0	1	0	0	961 298	3 886	957 412	-2 921
23	-	0	0	1	0	975 767	16 988		
24	-	0	0	0	1	1 002 625	42 479		
25	-	1	0	0	0	898 160	-63 354		
26	-	0	1	0	0	966 767	3 886		

$$T_t = (b_0 + \overline{a}) + b_1 t = 927330 + 1367\ t\,.$$

V předchozí tabulce jsou uvedeny hodnoty vstupních proměnných, odhadnutý model, odhadnuté sezónní faktory, odhadnuté hodnoty lineárního trendu a rezidua.

Na základě vypočítaných odhadů v tabulce můžeme model přepsat (v již odhadnutém tvaru)

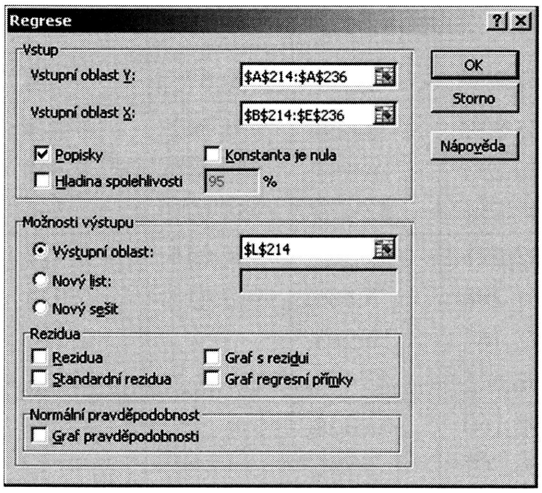
$$Y_t = 927330 + 1367\ t - 63\ 357 x_{1t} + 3\ 8886 x_{2t} + 16988 x_{3t} + 42479 x_{4t}\,.$$

V posledních čtyřech řádcích tabulky jsou údaje pro předpovědi HDP na další čtyři čtvrtletí. Mají hodnotu (v mil. Kč):

3. čtvrtletí 2012: $Y_{23} = 975\,767$,
4. čtvrtletí 2012: $Y_{24} = 1\,002\,625$,
1. čtvrtletí 2013: $Y_{25} = 898\,160$,
2. čtvrtletí 2013: $Y_{26} = 966\,767$.

Excel

Pro odhady parametrů v modelu $y_t = \beta_0 + \beta_1 t + \alpha_1 x_{1t} + \alpha_2 x_{2t} + \alpha_3 x_{3t} + \varepsilon_t$ využijeme proceduru **Regrese** z **Analýzy dat**.



	A	B	C	D	E
214	y_t	t	x_{1t}	x_{2t}	x_{3t}
215	843 399	1	1	0	0
216	905 917	2	0	1	0
217	935 100	3	0	0	1
218	978 157	4	0	0	0
219	889 080	5	1	0	0
220	970 995	6	0	1	0
221	997 237	7	0	0	1
222	991 099	8	0	0	0
223	888 452	9	1	0	0

Z výsledného výstupu z procedury **Regrese** je vidět, že odhady parametrů (až na zaokrouhlení) jsou stejné, jako v našich předchozích výpočtech. Tím jsme učinili první krok v řešení a dále postupujeme dle výše uvedených vzorců v souladu s teorií. Využijeme výstupu z regrese a přepočítáme hodnoty parametrů regresního modelu.

VÝSLEDEK

Regresní statistika	
Násobné R	0,92982
Hodnota spolehlivo	0,86457
Nastavená hodnota	0,8327
Chyba stř. hodnoty	18429,3
Pozorování	22

ANOVA

	Rozdíl	SS	MS	F	ýznamnost F
Regrese	4	3,7E+10	9,2E+09	27,13070457	3,5E-07
Rezidua	17	5,8E+09	3,4E+08		
Celkem	21	4,3E+10			

	Koeficienty	Chyba stř.	t Stat	Hodnota P	Dolní 95%	Horní 95%	Dolní 95,0'	Horní 95,0%
Hranice	969809	11113,3	87,2656	5,45776E-24	946362	993256,2044	946362	993256
t	1367,33	621,252	2,20093	0,041847868	56,6062	2678,061993	56,6062	2678,06
x1t	-105833	11176,8	-9,46896	3,42416E-08	-129413	-82251,56984	-129413	-82251,6
x2t	-38592,7	11159,5	-3,45828	0,00300374	-62137,2	-15048,19329	-62137,2	-15048,2
x3t	-25490,7	11672,3	-2,18386	0,043275029	-50117	-864,3244397	-50117	-864,324

	L	M
230	Hranice	969809,190909091
231	t	1367,33409090909
232	x1t	-105832,532575758
233	x2t	-38592,7
234	x3t	-25490,6659090909
235		
236		
237	$a_{průměr}$	=(M232+M233+M234)/4
238	S_{1+4j}	=M232-$M$237
239	S_{2+4j}	=M233-$M$237
240	S_{3+4j}	=M234-$M$237
241	S_{4+4j}	=-M237
242	Trend	
243	b_0	=M230+M237
244	b_1	=M231

Další výpočty jsou již zřejmé z následujícího výřezu tabulky.

	A	B	C	D	E	F	G	H	I	J
214	y_t	t	x_{1t}	x_{2t}	x_{3t}	x_{4t}	$Y_t = T_t + S_t$	S_t	T_t	$y_t - Y_t$
215	843399	1	1	0	0	0	=$M$243+$M$244*B215+$M$238*C215+$M$239*D215+$M$240*E215+$M$241*F215	=M238	=G215-H215	=A215-G215
216	905917	2	0	1	0	0	=$M$243+$M$244*B216+$M$238*C216+$M$239*D216+$M$240*E216+$M$241*F216	=M239	=G216-H216	=A216-G216
217	935100	3	0	0	1	0	=$M$243+$M$244*B217+$M$238*C217+$M$239*D217+$M$240*E217+$M$241*F217	=M240	=G217-H217	=A217-G217
218	978157	4	0	0	0	1	=$M$243+$M$244*B218+$M$238*C218+$M$239*D218+$M$240*E218+$M$241*F218	=M241	=G218-H218	=A218-G218
219	889080	5	1	0	0	0	=$M$243+$M$244*B219+$M$238*C219+$M$239*D219+$M$240*E219+$M$241*F219	-63353,557	=G219-H219	=A219-G219

■

7.3.2 Centrované klouzavé průměry

Příklad 7.9

V následující tabulce jsou uvedeny hodnoty měsíční časové řady počtu vydaných bytových stavebních povolení celkem. Aplikací centrovaných klouzavých průměrů vhodné délky odstraňte z časové řady sezónnost.

rok/měsíc	1	2	3	4	5	6	7	8	9	10	11	12
2002	2 571	2 725	3 742	4 127	4 705	4 656	4 619	4 097	4 263	4 089	3 397	2 970
2003	2 397	2 814	3 990	4 364	5 002	5 254	5 150	5 106	5 371	4 559	4 380	3 561
2004	3 085	3 046	4 333	4 741	4 988	5 207	4 430	4 900	4 882	4 333	3 989	3 530
2005	2 687	2 802	3 941	3 962	4 874	4 797	4 170	4 802	4 605	4 120	3 588	3 626
2006	2 802	2 640	3 453	3 906	4 937	4 804	4 621	5 281	4 871	4 683	3 966	3 813
2007	3 236	3 318	4 019	3 895	4 655	4 013	4 268	4 757	4 222	4 233	3 601	3 081
2008	3 001	3 228	3 945	4 217	4 686	4 642	4 410	4 318	4 283	3 999	3 313	3 347
2009	3 014	2 718	3 548	3 739	3 987	3 986	3 941	3 931	3 862	3 400	3 027	2 801
2010	2 297	2 608	3 254	3 472	3 833	3 494	3 471	4 142	3 476	3 298	3 014	2 799
2011	2 230	2 572	3 339	3 475	4 222	3 756	3 498	3 926	3 528	3 483	2 903	2 724
2012	2 279	2 244	2 804	3 002	3 394	3 119	3 138	3 176				

Zdroj: http://www.czso.cz/

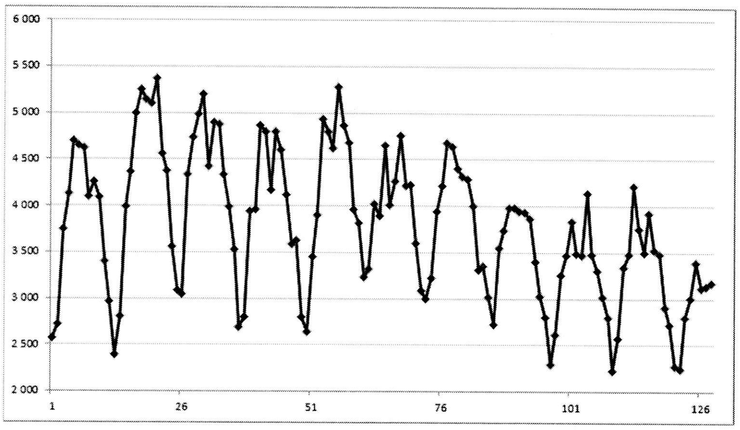
Řešení:

Podívejme se nejprve na průběh řady. Je zjevné, že se jedná o sezónní časovou řadu s délkou sezónnosti 12. Centrované klouzavé průměry by měly mít délku o jedničku větší, než je délka sezóny, přičemž krajní pozorování mají poloviční váhy. Jedná se tedy o vážené klouzavé průměry s váhami

$$\frac{1}{24}(1,2,2,2,2,2,2,2,2,2,2,2,1).$$

Potom první vyrovnaná hodnota existuje pro $t=7$ a je rovna

$$\overline{y}_7 = \frac{y_1 + 2y_2 + 2y_3 + \ldots + 2y_{11} + 2y_{12} + y_{13}}{24} =$$
$$= \frac{2571 + 2\cdot 2725 + 2\cdot 3742 + \ldots + 2\cdot 3397 + 2\cdot 2970 + 2397}{24} = 3823.$$



Druhá vyrovnaná hodnota pro $t = 8$ je rovna

$$\overline{y}_8 = \frac{y_2 + 2y_3 + 2y_4 + ... + 2y_{12} + 2y_{13} + y_{14}}{24} =$$

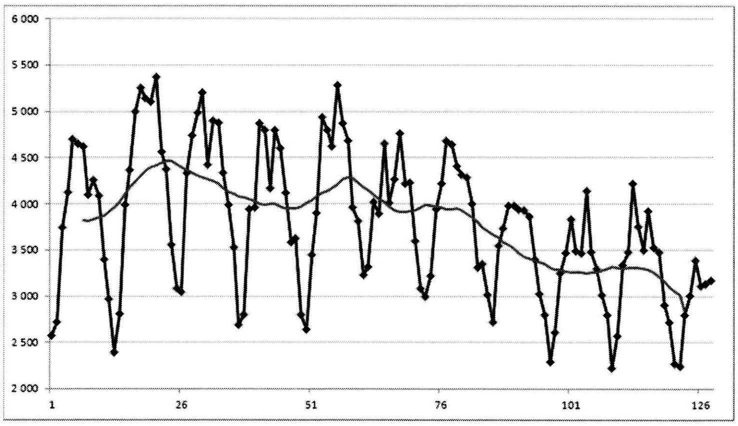
$$= \frac{2725 + 2 \cdot 3742 + 2 \cdot 4127 ... + 2 \cdot 2970 + +2 \cdot 2397 + 2814}{24} = 3819,$$

atd.

V další tabulce jsou uvedeny veškeré hodnoty spočítaných klouzavých průměrů, zaokrouhlené na celá čísla.

rok/měsíc	1	2	3	4	5	6	7	8	9	10	11	12
2002	-	-	-	-	-	-	3 823	3 819	3 833	3 854	3 876	3 913
2003	3 960	4 024	4 113	4 178	4 239	4 304	4 358	4 396	4 420	4 450	4 465	4 463
2004	4 431	4 392	4 363	4 333	4 308	4 290	4 272	4 245	4 219	4 170	4 133	4 111
2005	4 083	4 068	4 053	4 032	4 007	3 994	4 003	4 001	3 974	3 951	3 951	3 954
2006	3 973	4 012	4 043	4 078	4 117	4 140	4 166	4 213	4 264	4 287	4 275	4 231
2007	4 183	4 146	4 097	4 052	4 018	3 972	3 932	3 918	3 911	3 922	3 936	3 964
2008	3 996	3 984	3 968	3 961	3 939	3 938	3 950	3 929	3 891	3 855	3 806	3 749
2009	3 702	3 667	3 633	3 590	3 554	3 519	3 466	3 432	3 415	3 392	3 374	3 347
2010	3 307	3 296	3 289	3 269	3 264	3 263	3 260	3 256	3 258	3 262	3 278	3 305
2011	3 317	3 309	3 303	3 312	3 316	3 308	3 307	3 295	3 259	3 217	3 163	3 102
2012	3 060	3 014	2 836	-	-	-	-	-				

Je zřejmé, že při aplikaci klouzavých průměrů délky $k = 2m+1$, ztrácíme celkem $2m$ hodnot (m hodnot na začátku a m hodnot na konci časové řady), našem případě tedy 6 hodnot na začátku a 6 hodnot na konci časové řady .
Na obrázku je vidět, že centrované klouzavé průměry skutečně odstraní z časové řady sezónnost.



Excel

	A	B	C	D
1	rok	měsíc	y_t	centr. prům.
2	2002	1	2 571	
3		2	2 725	
4		3	3 742	
5		4	4 127	
6		5	4 705	
7		6	4 656	
8		7	4 619	3 823
9		8	4 097	3 819
10		9	4 263	3 833
11		10	4 089	3 854
12		11	3 397	3 876
13		12	2 970	3 913
14	2003	1	2 397	3 960
15		2	2 814	4 024
16		3	3 990	4 113
17		4	4 364	4 178
18		5	5 002	4 239
19		6	5 254	4 304

	A	B	C	D
1	rok	měsíc	y_t	centr. prům.
2	2002	1	2571	
3		2	2725	
4		3	3742	
5		4	4127	
6		5	4705	
7		6	4656	
8		7	4619	=(C2+2*SUMA(C3:C13)+C14)/24
9		8	4097	=(C3+2*SUMA(C4:C14)+C15)/24
10		9	4263	=(C4+2*SUMA(C5:C15)+C16)/24
11		10	4089	=(C5+2*SUMA(C6:C16)+C17)/24
12		11	3397	=(C6+2*SUMA(C7:C17)+C18)/24
13		12	2970	=(C7+2*SUMA(C8:C18)+C19)/24
14	2003	1	2397	=(C8+2*SUMA(C9:C19)+C20)/24
15		2	2814	=(C9+2*SUMA(C10:C20)+C21)/24
16		3	3990	=(C10+2*SUMA(C11:C21)+C22)/24
17		4	4364	=(C11+2*SUMA(C12:C22)+C23)/24
18		5	5002	=(C12+2*SUMA(C13:C23)+C24)/24
19		6	5254	=(C13+2*SUMA(C14:C24)+C25)/24

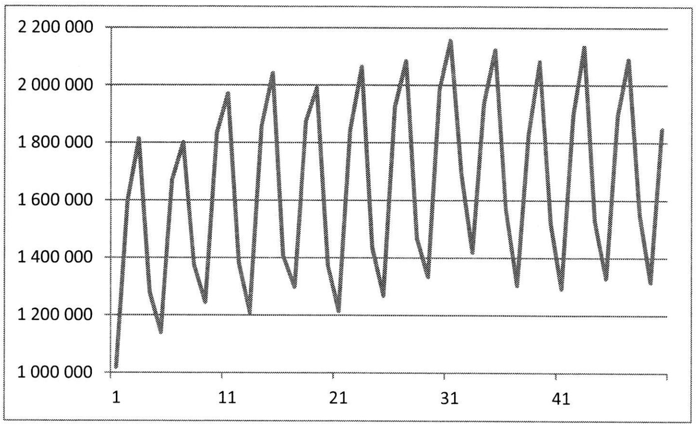
7.3.3 Sezónní dekompozice

Příklad 7.10
Máme k dispozici údaje o návštěvnosti v lázeňských ubytovacích zařízeních v ČR – počet přenocování. Jedná se o čtvrtletní časovou řadu za roky 2000-2012. Podívejme se na graf této řady.

rok/čtvrtletí	1	2	3	4
2000	1 018 861	1 599 177	1 813 359	1 278 351
2001	1 139 493	1 667 687	1 801 029	1 374 876
2002	1 244 772	1 833 449	1 970 930	1 381 884
2003	1 208 832	1 853 652	2 041 514	1 405 170
2004	1 298 295	1 873 335	1 990 663	1 375 765
2005	1 214 644	1 843 764	2 063 133	1 435 779
2006	1 267 630	1 925 980	2 084 536	1 465 747
2007	1 333 721	1 984 463	2 155 245	1 690 289
2008	1 418 724	1 936 674	2 122 679	1 567 543
2009	1 304 134	1 828 078	2 082 263	1 520 896
2010	1 292 444	1 896 746	2 133 186	1 528 006

rok/čtvrtletí	1	2	3	4
2011	1 328 323	1 892 751	2 088 813	1 547 837
2012	1 315 844	1 847 241		

Zdroj: http://www.czso.cz/



Je zjevné, že řada vykazuje sezónnost. Z obrázku lze usuzovat na aditivní model ve tvaru

$$y_t = T_t + S_t + C_t + \varepsilon_t, \quad t = 1, \dots, n.$$

Nejprve očistíme časovou řadu od sezónnosti. Veškeré výpočty budou patrné z následující tabulky. Sezónní složku odstraníme tak, že na časovou řadu budeme aplikovat centrované klouzavé průměry (viz předchozí příklad). Protože pracujeme se čtvrtletní časovou řadou, budou mít tyto průměry délku $k = 2m+1 = 5$ a váhy

$$\frac{1}{8}(1, 2, 2, 2, 1).$$

Tím získáme hodnoty řady, která bude obsahovat pouze trendovou a cyklickou složku $y'_t = T_t + C_t$ (5. sloupec tabulky). Přijdeme bohužel o 2 pozorování na začátku a 2 pozorování na konci časové řady ($m = 2$), která zůstanou nevyrovnaná. Pokud od původních hodnot řady odečteme hodnoty klouzavých průměrů, obdržíme sezónní a náhodnou složku (6. sloupec tabulky)

$$y_t - y'_t = T_t + S_t + C_t + \varepsilon_t - (T_t - C_t) = S_t + \varepsilon_t, \quad t = 1, \dots, n.$$

Nyní vezmeme z hodnot $S_t + \varepsilon_t$ údaje jen za všechna 1. čtvrtletí a spočítáme z nich průměr. Obdržíme tedy

$$S_t prum_1 = \frac{-333\,688{,}3 - 339\,997{,}1 - 403\,815{,}5 + \ldots + (-386\,696{,}9)}{11} = -381911{,}7\,.$$

Stejně postupujeme s údaji pro všechna 2. čtvrtletí

$$S_t prum_2 = \frac{183\,981{,}4 + 226\,566{,}3 + 229\,270{,}8 + \ldots + 180\,798{,}9}{11} = -201\,492{,}8$$

a rovněž pro 3. a 4. čtvrtletí. Za každá čtvrtletí sice máme jiný počet pozorování, ale tím, že pracujeme s průměrnými hodnotami, tento fakt nijak nevadí. Jen je třeba při výpočtu průměru respektovat počet čtvrtletí (tedy počet čísel, ze kterých průměr počítáme) – v datech máme 11 prvních, 11 druhých, 12 třetích a 12 čtvrtých čtvrtletí. Důvod, proč jsme počítali jednotlivé průměry ze součtu sezónní a náhodné složky pro jednotlivá čtvrtletí je ten, že průměrování těchto hodnot by mělo potlačit vliv náhodné složky (její vliv je v průměru nulový). Vypočtené hodnoty průměrů jsou v prvních čtyřech číselných řádcích v 7. sloupci tabulky a odpovídají průměrným sezónním odchylkám.

Na sezónní odchylky se někdy klade požadavek jejich vykompenzování v rámci periody (zde tedy v rámci jednoho roku)

$$S_{t1} + S_{t2} + S_{t3} + S_{t4} = 0\,.$$

Dosáhneme toho tak, že spočítané průměry sečteme (5. číselný řádek v 7. sloupci)

$$S_t prum = S_t prum_1 + S_t prum_2 + S_t prum_3 + S_t prum_4 =$$
$$= -381\,911{,}6 + 201492{,}8 + 377412{,}2 - 192865{,}9 = 4\,127{,}5$$

a dále provedeme vlastní výpočet sezónních odchylek S_{ti}, $i = 1,2,3,4$ dle vzorce

$$S_{ti} = S_t prum_i - \frac{S_t prum}{4}, \quad i = 1,2,3,4\,.$$

Po dosazení tedy obdržíme

$$S_{t1} = S_t prum_1 - \frac{S_t prum}{4} = -381\,911{,}6 - \frac{4127{,}5}{4} = -382\,943{,}5\,,$$

$$S_{t2} = S_t prum_2 - \frac{S_t prum}{4} = 201492{,}8 - \frac{4127{,}5}{4} = 200\ 461{,}0\,,$$

atd., viz 8. sloupec tabulky.

Sezónně očištěné hodnoty časové řady získáme nakonec tak, že od původních napozorovaných hodnot odečteme po jednotlivých čtvrtletích hodnoty sezónních odchylek (9. sloupec tabulky). Při výpočtu hlídáme, abychom od pozorování v čase t v určitém čtvrtletí odečetli sezónní odchylku pro stejné čtvrtletí.

$$y_t adj = y_t - S_{ti}, \quad i = 1,2,3,4\,, \quad t = 1,\ldots,n\,.$$

t	rok		y_t	$T_t + C_t$	$S_t + \varepsilon_t$	$S_t prum_i$	S_{ti}	$y_t adj = y_t - S_{ti}$	E_t
1	2000	Q1	1 018 861			-381911,7	-382 943,5	1 401 804,5	
2		Q2	1 599 177			201492,8	200 461,0	1 398 716,0	
3		Q3	1 813 359	1 442 516	370 843,0	377412,2	376 380,4	1 436 978,6	-5 537,4
4		Q4	1 278 351	1 466 159	-187 807,8	-192865,9	-193 897,8	1 472 248,8	6 090,1
5	2001	Q1	1 139 493	1 473 181	-333 688,3	4 127,5	-382 943,5	1 522 436,5	49 255,3
6		Q2	1 667 687	1 483 706	183 981,4		200 461,0	1 467 226,0	-16 479,6
7		Q3	1 801 029	1 508 931	292 097,9		376 380,4	1 424 648,6	-84 282,5
8		Q4	1 374 876	1 542 811	-167 935,3		-193 897,8	1 568 773,8	25 962,6
9	2002	Q1	1 244 772	1 584 769	-339 997,1		-382 943,5	1 627 715,5	42 946,4
10		Q2	1 833 449	1 606 883	226 566,3		200 461,0	1 632 988,0	26 105,3
11		Q3	1 970 930	1 603 266	367 663,8		376 380,4	1 594 549,6	-8 716,6
12		Q4	1 381 884	1 601 299	-219 415,1		-193 897,8	1 575 781,8	-25 517,3
13	2003	Q1	1 208 832	1 612 648	-403 815,5		-382 943,5	1 591 775,5	-20 872,0
14		Q2	1 853 652	1 624 381	229 270,8		200 461,0	1 653 191,0	28 809,8
15		Q3	2 041 514	1 638 475	403 039,1		376 380,4	1 665 133,6	26 658,8
16		Q4	1 405 170	1 652 118	-246 948,1		-193 897,8	1 599 067,8	-53 050,3
17	2004	Q1	1 298 295	1 648 222	-349 927,1		-382 943,5	1 681 238,5	33 016,4
18		Q2	1 873 335	1 638 190	235 144,9		200 461,0	1 672 874,0	34 683,9
19		Q3	1 990 663	1 624 058	366 604,9		376 380,4	1 614 282,6	-9 775,5
20		Q4	1 375 765	1 609 905	-234 140,4		-193 897,8	1 569 662,8	-40 242,6
21	2005	Q1	1 214 644	1 615 268	-400 623,8		-382 943,5	1 597 587,5	-17 680,2
22		Q2	1 843 764	1 631 828	211 935,8		200 461,0	1 643 303,0	11 474,8
23		Q3	2 063 133	1 645 953	417 179,8		376 380,4	1 686 752,6	40 799,4
24		Q4	1 435 779	1 662 854	-227 074,5		-193 897,8	1 629 676,8	-33 176,7
25	2006	Q1	1 267 630	1 675 806	-408 175,9		-382 943,5	1 650 573,5	-25 232,4
26		Q2	1 925 980	1 682 227	243 752,8		200 461,0	1 725 519,0	43 291,8
27		Q3	2 084 536	1 694 235	390 301,4		376 380,4	1 708 155,6	13 921,0

t	rok		y_t	$T_t + C_t$	$S_t + \varepsilon_t$	$S_t prum_i$	S_{ti}	$y_t adj = y_t - S_{ti}$	E_t
28		Q4	1 465 747	1 709 806	-244 059,4		-193 897,8	1 659 644,8	-50 161,6
29	2007	Q1	1 333 721	1 725 955	-392 234,4		-382 943,5	1 716 664,5	-9 290,9
30		Q2	1 984 463	1 762 862	221 601,3		200 461,0	1 784 002,0	21 140,3
31		Q3	2 155 245	1 801 555	353 690,1		376 380,4	1 778 864,6	-22 690,2
32		Q4	1 690 289	1 806 207	-115 917,6		-193 897,8	1 884 186,8	77 980,2
33	2008	Q1	1 418 724	1 796 162	-377 438,3		-382 943,5	1 801 667,5	5 505,3
34		Q2	1 936 674	1 776 748	159 925,8		200 461,0	1 736 213,0	-40 535,2
35		Q3	2 122 679	1 747 081	375 597,8		376 380,4	1 746 298,6	-782,6
36		Q4	1 567 543	1 719 183	-151 640,0		-193 897,8	1 761 440,8	42 257,8
37	2009	Q1	1 304 134	1 700 557	-396 422,5		-382 943,5	1 687 077,5	-13 479,0
38		Q2	1 828 078	1 689 674	138 404,4		200 461,0	1 627 617,0	-62 056,6
39		Q3	2 082 263	1 682 382	399 881,5		376 380,4	1 705 882,6	23 501,1
40		Q4	1 520 896	1 689 504	-168 607,8		-193 897,8	1 714 793,8	25 290,1
41	2010	Q1	1 292 444	1 704 453	-412 008,6		-382 943,5	1 675 387,5	-29 065,1
42		Q2	1 896 746	1 711 707	185 039,3		200 461,0	1 696 285,0	-15 421,7
43		Q3	2 133 186	1 717 080	416 105,6		376 380,4	1 756 805,6	39 725,3
44		Q4	1 528 006	1 721 066	-193 059,9		-193 897,8	1 721 903,8	837,9
45	2011	Q1	1 328 323	1 715 020	-386 696,9		-382 943,5	1 711 266,5	-3 753,4
46		Q2	1 892 751	1 711 952	180 798,9		200 461,0	1 692 290,0	-19 662,1
47		Q3	2 088 813	1 712 871	375 941,9		376 380,4	1 712 432,6	-438,5
48		Q4	1 547 837	1 705 623	-157 785,5		-193 897,8	1 741 734,8	36 112,3
49	2012	Q1	1 315 844	-	-	-	-382 943,5	1 698 787,5	-
50		Q2	1 847 241	-	-	-	200 461,0	1 646 780,0	-

V posledním sloupci jsou hodnoty reziduální složky $E_t = T_t + C_t - y_t adj$. S takto očištěnou časovou řadou již můžeme dále pracovat např. pomocí trendové analýzy.

Excel

	A	B	C	D	E	F	G	H	I
1	t	čtvrtletí	y_t	$T_t + C_t$	$S_t + \varepsilon_t$	$S_t + prum_i$	S_{ti}	E_t	$y_t adj = y_t - S_{ti}$
2	1	1	1018861			=G2-$G$6/4	=AVERAGEIF($B$4:$B$49;1;$E$4:$E$49)		=C2-F2
3	2	2	1599177			=G3-$G$6/4	=AVERAGEIF($B$4:$B$49;2;$E$4:$E$49)		=C3-F3
4	3	3	1813359	=1/8*(C2+2*C3+2*C4+2*C5+C6)	=C4-D4	=G4-$G$6/4	=AVERAGEIF($B$4:$B$49;3;$E$4:$E$49)	=I4-D4	=C4-F4
5	4	4	1278351	=1/8*(C3+2*C4+2*C5+2*C6+C7)	=C5-D5	=G5-$G$6/4	=AVERAGEIF($B$4:$B$49;4;$E$4:$E$49)	=I5-D5	=C5-F5
6	5	1	1139493	=1/8*(C4+2*C5+2*C6+2*C7+C8)	=C6-D6	-382943,524857955	**=SUMA(G2:G5)**	=I6-D6	=C6-F6
7	6	2	1667687	=1/8*(C5+2*C6+2*C7+2*C8+C9)	=C7-D7	200460,975142045		=I7-D7	=C7-F7
8	7	3	1801029	=1/8*(C6+2*C7+2*C8+2*C9+C10)	=C8-D8	376380,352982955	Countif	=I8-D8	=C8-F8
9	8	4	1374876	=1/8*(C7+2*C8+2*C9+2*C10+C11)	=C9-D9	-193897,803267045	=COUNTIF($B$4:$B$49;1)	=I9-D9	=C9-F9
10	9	1	1244772	=1/8*(C8+2*C9+2*C10+2*C11+C12)	=C10-D10	-382943,524857955	=COUNTIF($B$4:$B$49;2)	=I10-D10	=C10-F10
11	10	2	1833449	=1/8*(C9+2*C10+2*C11+2*C12+C13)	=C11-D11	200460,975142045	=COUNTIF($B$4:$B$49;3)	=I11-D11	=C11-F11
12	11	3	1970930	=1/8*(C10+2*C11+2*C12+2*C13+C14)	=C12-D12	376380,352982955	=COUNTIF($B$4:$B$49;4)	=I12-D12	=C12-F12
13	12	4	1381884	=1/8*(C11+2*C12+2*C13+2*C14+C15)	=C13-D13	-193897,803267045		=I13-D13	=C13-F13
14	13	1	1208832	=1/8*(C12+2*C13+2*C14+2*C15+C16)	=C14-D14	-382943,524857955		=I14-D14	=C14-F14

	A	B	C	D	E	F	G	H	I
1	t	čtvrtletí	y_t	$T_t + C_t$	$S_t + \varepsilon_t$	S_t + $prum_i$	S_{ti}	E_t	$y_t\,adj = y_t - S_{ti}$
2	1	1	1 018 861			-382 943,5	-381911,7		1 401 804,5
3	2	2	1 599 177			200 461,0	201492,8		1 398 716,0
4	3	3	1 813 359	1 442 516	370 843,0	376 380,4	377412,2	-5 537,4	1 436 978,6
5	4	4	1 278 351	1 466 159	-187 807,8	-193 897,8	-192865,9	6 090,1	1 472 248,8
6	5	1	1 139 493	1 473 181	-333 688,3	-382 943,5	**4 127,5**	49 255,3	1 522 436,5
7	6	2	1 667 687	1 483 706	183 981,4	200 461,0		-16 479,6	1 467 226,0
8	7	3	1 801 029	1 508 931	292 097,9	376 380,4	Countif	-84 282,5	1 424 648,6
9	8	4	1 374 876	1 542 811	-167 935,3	-193 897,8	11	25 962,6	1 568 773,8
10	9	1	1 244 772	1 584 769	-339 997,1	-382 943,5	11	42 946,4	1 627 715,5
11	10	2	1 833 449	1 606 883	226 566,3	200 461,0	12	26 105,3	1 632 988,0
12	11	3	1 970 930	1 603 266	367 663,8	376 380,4	12	-8 716,6	1 594 549,6
13	12	4	1 381 884	1 601 299	-219 415,1	-193 897,8		-25 517,3	1 575 781,8
14	13	1	1 208 832	1 612 648	-403 815,5	-382 943,5		-20 872,0	1 591 775,5

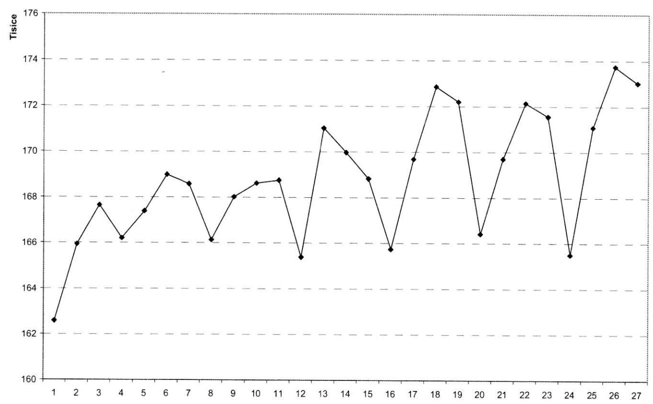
▫

Příklad 7.11

V následující tabulce jsou údaje o počtu pracovníků v oblasti ubytování a stravování. Jedná se čtvrtletí za roky 2000 - 2006. Podívejme se na graf této řady.

rok/čtvrtletí	1	2	3	4
2000	162 590	165 930	167 639	166 200
2001	167 378	168 976	168 572	166 130
2002	168 004	168 603	168 740	165 378
2003	171 025	169 958	168 814	165 737
2004	169 674	172 855	172 196	166 404
2005	169 690	172 141	171 544	165 498
2006	171 075	173 737	173 022	

Zdroj: http://www.czso.cz/



Je zjevné, že řada vykazuje sezónnost. Z obrázku lze usuzovat na multiplikativní model ve tvaru

$$y_t = T_t \cdot S_t \cdot C_t \cdot \varepsilon_t, \quad t = 1,\ldots,n\,.$$

Budeme postupovat analogicky, jako v předchozím příkladu. Nejprve očistíme časovou řadu od sezónnosti. Veškeré výpočty budou patrné z uvedené tabulky. Sezónní složku odstraníme tak, že na časovou řadu budeme aplikovat centrované klouzavé průměry (viz předchozí příklad). Protože pracujeme se čtvrtletní časovou řadou, budou mít tyto průměry délku k=2m+1=5 a váhy

$$\frac{1}{8}(1,2,2,2,1)\,.$$

Tím získáme hodnoty řady, které budou obsahovat pouze trendovou a cyklickou složku $y'_t = T_t \cdot C_t$ (5. sloupec tabulky). Přijdeme bohužel o 2 pozorování na začátku a dvě pozorování na konci časové řady ($m = 2$), která zůstanou nevyrovnaná. Pokud původní hodnoty řady vydělíme hodnotou klouzavých průměrů, obdržíme sezónní indexy, zahrnující náhodnou složku (6. sloupec tabulky)

$$\frac{y_t}{y'_t} = \frac{T_t \cdot S_t \cdot C_t \cdot \varepsilon_t}{T_t \cdot C_t} = S_t \cdot \varepsilon_t, \quad t = 1,\ldots,n\,.$$

Nyní vezmeme z hodnot $S_t \cdot \varepsilon_t$ údaje za všechna 1. čtvrtletí a spočítáme z nich průměr. Obdržíme tedy průměrný sezónní index

$$S_t prum_1 = \frac{0,998288 + 1,000928 + 1,013273 + 0,999893 + 0,998022 + 1,0025}{6} =$$
$$= 1,002151.$$

Stejně postupujeme pro všechna 2. čtvrtletí

$$S_t prum_2 = \frac{1,007172 + 1,004933 + 1,006633 + 1,015607 + 1,0136}{5} = 1,009589$$

a rovněž pro 3. a 4. čtvrtletí. Za každá čtvrtletí sice máme jiný počet pozorování a proto pracujeme s průměrnými hodnotami, čímž tento fakt nijak nevadí. Důvod, proč jsme počítali jednotlivé průměry ze součinu sezónní a náhodné složky pro jednotlivá čtvrtletí je opět ten, že průměrování těchto hodnot by mělo potlačit vliv náhodné složky (její vliv je v průměru nulový). Vypočtené hodnoty průměrů jsou v prvních čtyřech číselných řádcích v 7. sloupci tabulky. Poté spočítané průměry sečteme (5. číselný řádek v 7. sloupci)

$$S_t prum = S_t prum_1 + S_t prum_2 + S_t prum_3 + S_t prum_4 =$$
$$= 1,002151 + 1,009589 + 1,006446 + 0,982604 = 4,000790$$

a spočítáme sezónní indexy S_{ti}, $i = 1,2,3,4$. Při výpočtu těchto indexů vycházíme z požadavku

$$S_{t1} + S_{t2} + S_{t3} + S_{t4} = 4\,.$$

Výše uvedený předpoklad má opět zaručit jednoznačnost dekompozičního rozkladu – tedy aby se vliv sezónních faktorů v rámci jednoho kalendářního roku celkově vykompenzoval.

Vlastní výpočet sezónních indexů S_{ti}, $i = 1,2,3,4$ se provádí dle vzorce

$$S_{ti} = \frac{4}{S_t prum} \cdot S_t prum_i\,, \quad i = 1,2,3,4\,.$$

Po dosazení tedy obdržíme

$$S_{t1} = \frac{4}{S_t prum} \cdot S_t prum_1 = \frac{4}{4,000790} \cdot 1,002151 = 1,001953\,,$$

$$S_{t2} = \frac{4}{S_t prum} \cdot S_t prum_2 = \frac{4}{4,000790} \cdot 1,009589 = 1,009589\,,$$

atd., viz 8. sloupec tabulky.

t	rok		y_t	$T_t \cdot C_t$	$S_t \cdot \varepsilon_t$	$S_t prum_i$	S_{ti}	y_t / S_{ti}	E_t
1	2000	Q1	162 590			1,002151	1,001953	162 274	
2		Q2	165 930			1,009589	1,009390	164 386	
3		Q3	167 639	166 188	1,008731	1,006446	1,006247	166 599	1,002469
4		Q4	166 200	167 168	0,994213	0,982604	0,982410	169 176	1,012014
5	2001	Q1	167 378	167 665	0,998288	**4,000790**	1,001953	167 052	0,996343
6		Q2	168 976	167 773	1,007172		1,009390	167 404	0,997803
7		Q3	168 572	167 843	1,004349		1,006247	167 526	0,998114
8		Q4	166 130	167 874	0,989611		0,982410	169 105	1,00733

t	rok		y_t	$T_t \cdot C_t$	$S_t \cdot \varepsilon_t$	$S_t prum_i$	S_{ti}	y_t / S_{ti}	E_t
9	2002	Q1	168 004	167 848	1,000928		1,001953	167 677	0,998977
10		Q2	168 603	167 775	1,004933		1,009390	167 035	0,995585
11		Q3	168 740	168 059	1,004054		1,006247	167 692	0,99782
12		Q4	165 378	168 606	0,980855		0,982410	168 339	0,998417
13	2003	Q1	171 025	168 784	1,013273		1,001953	170 691	1,011298
14		Q2	169 958	168 838	1,006633		1,009390	168 377	0,997269
15		Q3	168 814	168 715	1,000587		1,006247	167 765	0,994375
16		Q4	165 737	168 908	0,981225		0,982410	168 704	0,998794
17	2004	Q1	169 674	169 693	0,999893		1,001953	169 344	0,997944
18		Q2	172 855	170 199	1,015607		1,009390	171 247	1,006159
19		Q3	172 196	170 284	1,011225		1,006247	171 126	1,004946
20		Q4	166 404	170 197	0,977713		0,982410	169 383	0,995219
21	2005	Q1	169 690	170 026	0,998022		1,001953	169 359	0,996077
22		Q2	172 141	169 832	1,0136		1,009390	170 540	1,004171
23		Q3	171 544	169 892	1,009729		1,006247	170 479	1,00346
24		Q4	165 498	170 264	0,972007		0,982410	168 461	0,98941
25	2006	Q1	171 075	170 648	1,0025		1,001953	170 742	1,000546
26		Q2	173 737				1,009390	172 121	
27		Q3	173 022				1,006247	171 948	

Sezónně očištěné hodnoty časové řady získáme nakonec tak, že původní napozorované hodnoty vydělíme po jednotlivých čtvrtletích hodnotami sezónních indexů (9. sloupec tabulky). Při výpočtu hlídáme, abychom pozorování v čase t v určitém čtvrtletí dělili sezónním indexem pro stejné čtvrtletí.

$$y_t adj = \frac{y_t}{S_{ti}}, \quad i = 1,2,3,4; \ t = 1,\dots,n .$$

V posledním sloupci tabulky jsou hodnoty reziduální složky. S takto očištěnou časovou řadou již můžeme dále pracovat např. pomocí trendové analýzy.

Excel

Při výpočtech v Excelu budeme postupovat obdobným způsobem, jako v předchozím příkladu v případě modelu s aditivní dekompozicí. Použijeme výše uvedené vzorce s respektováním multiplikativního modelu, který s sebou pochopitelně nese jinou podobu použitých vzorců.

	A	B	C	D	E	F	G	H	I
1	t	čtvrtletí	y_t	$T_t * C_t$	$S_t * \varepsilon_t$	S_t * $prum_i$	S_{ti}	y_t / S_{ti}	E_t
2	1	1	162 590			1,001953	1,002150671	162 274	
3	2	2	165 930			1,009390	1,009589131	164 386	
4	3	3	167 639	166 188	1,008731438	1,006247	1,006445826	166 599	1,002468805
5	4	4	166 200	167 168	0,994212715	0,982410	0,982603903	169 176	1,01201405
6	5	1	167 378	167 665	0,998288279	1,001953	**4,000789531**	167 052	0,99634252
7	6	2	168 976	167 773	1,007172037	1,009390		167 404	0,997802773
8	7	3	168 572	167 843	1,004349245	1,006247		167 526	0,998113818
9	8	4	166 130	167 874	0,989611273	0,982410		169 105	1,00733022
10	9	1	168 004	167 848	1,0009277	1,001953		167 677	0,998976796
11	10	2	168 603	167 775	1,004933399	1,009390		167 035	0,995584961
12	11	3	168 740	168 059	1,004053576	1,006247		167 692	0,997819985
13	12	4	165 378	168 606	0,980855038	0,982410		168 339	0,998417204
14	13	1	171 025	168 784	1,013273059	1,001953		170 691	1,011298092

	A	B	C	D	E	F	G	H	I
1	t	čtvrtletí	y_t	$T_t * C_t$	$S_t * \varepsilon_t$	S_t * $prum_i$	S_{ti}	y_t / S_{ti}	E_t
2	1	1	162590,499789684			=4/$G$6*G2	=PRŮMĚR(E6;E10;E14;E18;E22;E26)	=C2/F2	
3	2	2	165929,528746906			=4/$G$6*G3	=PRŮMĚR(E7;E11;E15;E19;E23)	=C3/F3	
4	3	3	167639,463697706	=1/8*(C2+2*C3+2*C4+2*C5+C6)	=C4/D4	=4/$G$6*G4	=PRŮMĚR(E8;E12;E16;E4;E20;E24)	=C4/F4	=H4/D4
5	4	4	166200,262340344	=1/8*(C3+2*C4+2*C5+2*C6+C7)	=C5/D5	=4/$G$6*G5	=PRŮMĚR(E9;E13;E5;E17;E21;E25)	=C5/F5	=H5/D5
6	5	1	167378,190069682	=1/8*(C4+2*C5+2*C6+2*C7+C8)	=C6/D6	1,00195290221465	**=SUMA(G2:G5)**	=C6/F6	=H6/D6
7	6	2	168976,315048931	=1/8*(C5+2*C6+2*C7+2*C8+C9)	=C7/D7	1,00938989522214		=C7/F7	=H7/D7
8	7	3	168572,489932591	=1/8*(C6+2*C7+2*C8+2*C9+C10)	=C8/D8	1,00624721024329		=C8/F8	=H8/D8
9	8	4	166130,072833637	=1/8*(C7+2*C8+2*C9+2*C10+C11)	=C9/D9	0,982409992319921		=C9/F9	=H9/D9
10	9	1	168004,068190498	=1/8*(C8+2*C9+2*C10+2*C11+C12)	=C10/D10	1,00195290221465		=C10/F10	=H10/D10
11	10	2	168602,988478547	=1/8*(C9+2*C10+2*C11+2*C12+C13)	=C11/D11	1,00938989522214		=C11/F11	=H11/D11
12	11	3	168740,093001852	=1/8*(C10+2*C11+2*C12+2*C13+C14)	=C12/D12	1,00624721024329		=C12/F12	=H12/D12
13	12	4	165377,915785903	=1/8*(C11+2*C12+2*C13+2*C14+C15)	=C13/D13	0,982409992319921		=C13/F13	=H13/D13
14	13	1	171024,766038787	=1/8*(C12+2*C13+2*C14+2*C15+C16)	=C14/D14	1,00195290221465		=C14/F14	=H14/D14

KAPITOLA VIII

INDEXY A ABSOLUTNÍ ROZDÍLY

Zboží	$\frac{p_1}{p_0} = Ip$	$\frac{Q_1}{Q_0} = IQ$	$Q_1 - Q_0 = \Delta Q$	$\frac{\ln Ip}{\ln IQ}\Delta Q$
A	1,250	0,833	– 40	48,956
B	1,500	1,200	40	88,956
C	1,600	0,960	–10	115,135
D	1,000	0,667	–70	0
E	0,889	1,778	70	-14,330
Součet	×	×	–10	238,717

8

8.1

Výpo
pinar
vlože
a výp
pod č

Nejp
ství q

Výpo
příslu
Podo
rozkl

8 Indexy a absolutní rozdíly

8.1 Indexy bazické a řetězové

Výpočty indexů a rozdílů spočívají v jednoduchých opakujících se operacích se skupinami čísel (násobení a sčítání), které lze provádět v Excelu běžným způsobem, tedy vložením příslušného vzorce do pole, jeho zkopírováním do dalších polí ve sloupci a výpočtem součtu. V ukázce použijeme údaje z příkladu, zařazeného v této kapitole pod číslem 11.

Excel

	A	B	C	D
1	p0	p1	q0	q1
2	60	70	200	300
3	40	30	500	400
4	80	100	300	400

Nejprve provedeme některé dílčí výpočty (budeme potřebovat součiny cen p a množství q v různých obdobích). Dostáváme:

	Q0=p0q0	Q1=p1q1	p0q1	p1q0
	12000	21000	18000	14000
	20000	12000	16000	15000
	24000	40000	32000	30000
Součty	56000	73000	66000	59000

Výpočty indexů, případně diferencí, pak provádíme jednoduchým dosazováním do příslušných vzorců.
Podobně si můžeme rovněž připravit dílčí výpočty pro použití logaritmické metody rozkladu indexů apod.

ln(p1/p0)	ln(q1/q0)	ln(Q1/Q0)
0,154151	0,405465	0,559616
-0,28768	-0,22314	-0,51083
0,223144	0,287682	0,510826

Příklad 8.1

V tabulce je uvedena časová řada spotřeby piva v České republice (v litrech na osobu za rok) v letech 2003 až 2010. Charakterizujte vývoj spotřeby piva pomocí bazických indexů (1989 = 100) a řetězových indexů.

Rok	1989	2003	2004	2005	2006	2007	2008	2009	2010
Spotřeba	151,0	161,7	160,5	163,5	159,1	159,1	156,6	150,7	144,4

Řešení:

Chceme-li vývoj charakterizovat ve vztahu k roku 1989, použijeme bazické indexy (tj. indexy se stálým základem, spotřebou piva v roce 1989). Například pro rok 2003 dostaneme

$$I_{03/89} = \frac{161,7}{151,0} = 1,071.$$

To znamená, že v roce 2003 vzrostla spotřeba piva o 7,1 % vůči roku 1989.

Chceme-li posoudit vývoj vždy k předchozímu roku, použijeme řetězové indexy. Například pro rok 2004 ve vztahu k roku 2003 dostaneme

$$I_{04/03} = \frac{160,5}{161,7} = 0,993.$$

To znamená, že v roce 2004 poklesla spotřeba piva v ČR o 0,7 % proti předchozímu roku. Hodnoty bazických a řetězových indexů (v %) jsou uvedeny v následující tabulce:

Indexy	1989	2003	2004	2005	2006	2007	2008	2009	2010
bazické	100,0	107,1	106,3	108,3	105,4	105,4	103,7	99,8	95,6
řetězové	–	–	99,3	101,9	97,3	100,0	98,4	96,2	95,8

Průměrný koeficient růstu pro období 2004 – 2010 bude geometrickým průměrem spočtených řetězových indexů, tedy

$$\sqrt[7]{0,993 \cdot 1,019 \cdot 0,973 \cdot 1,000 \cdot 0,984 \cdot 0,962 \cdot 0,958} = 0,984.$$

Stejného výsledku dosáhneme (po úpravě vzorce použitého pro předchozí výpočet) s využitím hodnoty spotřeby v roce 2010 a 2004, případně s využitím příslušných bazických indexů, takto:

$$\sqrt[7]{144,4/161,7}=0,984\,;\ \sqrt[7]{95,6/107,1}=0,984\,.$$

Příklad 8.2

V tabulce jsou bazické indexy počtu dokončených bytů v ČR v letech 2004 až 2007 se základem v roce 2004 ($I_{i/04}$), a dále bazické indexy počtu dokončených bytů v letech 2007 až 2010 se základem v roce 2007 ($I_{i/07}$), obojí v procentech. Dopočítejte chybějící bazické indexy v obou řadách.

Rok	$I_{i/04}$	$I_{i/07}$
2004	100,0	.
2005	101,8	.
2006	93,6	.
2007	129,1	100,0
2008	.	92,2
2009	.	92,4
2010	.	87,5

Řešení:

Protože známe z první řady bazický index v roce 2007 a tento rok je základem srovnání v řadě druhé, můžeme provést přepočet takto:

Chybějící hodnoty bazických indexů se základem v roce 2004 získáme tak, že indexy se základem v roce 2007 budeme násobit podílem 129,1/100 (stanoveným z indexů roku 2007, které jsou zvýrazněny v následující tabulce):

Rok	$I_{i/04}$	$I_{i/07}$
2004	100,0	77,5
2005	101,8	78,9
2006	93,6	72,5
2007	**129,1**	**100,0**
2008	119,0	92,2
2009	119,3	92,4
2010	113,0	87,5

$$92{,}2 \cdot 129{,}1/100 = 119{,}0$$

$$92{,}4 \cdot 129{,}1/100 = 119{,}3$$

$$87{,}5 \cdot 129{,}1/100 = 113{,}0.$$

Při výpočtu chybějících bazických indexů se základem v roce 2007 budeme indexy se základem v roce 2004 podílem 129,1/100 naopak dělit (neboli násobit podílem 100/129,1):

$$100 \cdot 100/129{,}1 = 77{,}5$$

$$101{,}8 \cdot 100/129{,}1 = 78{,}9$$

$$93{,}6 \cdot 100/129{,}1 = 72{,}5\ .$$

□

8.2 Individuální indexy

8.2.1 Individuální indexy jednoduché

Příklad 8.3

Porovnejte rozlohu, počet obyvatel a hustotu obyvatelstva České republiky (ČR) a Slovenska (S) v letech 1996 a 2011, máte-li následující údaje (k 1. lednu):

	Počet obyvatel (tis.)		Rozloha (km^2)		Hustota (na km^2)	
Rok	ČR	S	ČR	S	ČR	S
1996	10 300	5 410	78 866	49 036	131	110
2011	10 533	5 435	78 865	49 037	134	111

Řešení:

Označíme-li jako Q počet obyvatel, jako q rozlohu státu a jako p hustotu obyvatel, dostaneme například pro rok 2011:

$$IQ = \frac{Q_1}{Q_0} = \frac{5435}{10533} = 0{,}516\ ;\ \Delta Q = Q_1 - Q_0 = 5\,435 - 10\ 533 = -\,5\,098\,.$$

V roce 2011 mělo Slovensko téměř o polovinu, tj. o cca 5 mil. obyvatel méně než Česká republika.

$$Iq = \frac{q_1}{q_0} = \frac{49\ 037}{78\ 865} = 0{,}622\ ;\ \Delta q = q_1 - q_0 = 49\ 037 - 78\ 865 = -29\ 828\ .$$

Rozloha Slovenska byla o 37,8 %, tj. o 29 828 km^2 menší než rozloha České republiky.

$$Ip = \frac{p_1}{p_0} = \frac{111}{134} = 0{,}828\,;\ \Delta p = p_1 - p_0 = 111 - 134 = -23\,.$$

Hustota obyvatelstva Slovenska byla o 17,2 %, tj. o 23 obyvatele na 1 km^2 nižší než hustota obyvatel České republiky.
Při srovnání veličin v roce 2011 a 1996 dostáváme pro Českou republiku:

$$IQ = \frac{Q_1}{Q_0} = \frac{10533}{10300} = 1{,}023\,;\ Ip = \frac{p_1}{p_0} = \frac{134}{131} = 1{,}023\,.$$

Počet obyvatel stejně jako hustota vzrostl za 15 let o 2,3 %, neboť rozloha státu je prakticky stejná.
Podobně pro Slovensko dostáváme u obou veličin nárůst o 0,5 % (výpočet pro Slovensko je nutno provést přesněji, aby nebyl nepříznivě ovlivněn zaokrouhlením):

$$IQ = \frac{Q_1}{Q_0} = \frac{5435}{5410} = 1{,}005\,;\ Ip = \frac{p_1}{p_0} = \frac{110{,}8}{110{,}3} = 1{,}005\,.$$

□

Příklad 8.4
Prášku na praní značky Nela se prodalo o 6 % méně, ale tržba z jeho prodeje byla o 2 % vyšší než u prášku značky Lena. Jaký je vztah mezi cenami prášků na praní těchto značek?

Řešení:
Cenu za kilogram prášku (p) lze vyjádřit jako podíl

$$p = \frac{Q}{q}\,,$$

kde Q značí tržbu, q množství prodaného prášku. Pro index ceny prášku na praní, bereme-li za základ cenu prášku Lena, platí

$$Ip = \frac{p_N}{p_L} = \frac{\dfrac{Q_N}{q_N}}{\dfrac{Q_L}{q_L}} = \frac{Q_N}{Q_L} / \frac{q_N}{q_L} = 1,02/0,94 = 1,085.$$

Index ceny je 1,085; prášek Nela je tedy o 8,5 % dražší než prášek Lena.

□

Příklad 8.5

Jak se změnilo prodané množství prášku na praní značky Nela v červnu oproti květnu sledovaného roku, zůstala-li tržba z prodeje tohoto prášku stejná, ale jeho cena vzrostla o 5 %?

Řešení:

$$Ip = \frac{p_1}{p_0} = \frac{\dfrac{Q_1}{q_1}}{\dfrac{Q_0}{q_0}} = \frac{Q_1}{Q_0} / \frac{q_1}{q_0},$$

$$1,05 = 1/\frac{q_1}{q_0},$$

$$Iq = 0,952.$$

Nezměnila-li se tržba z prodeje prášku, znamená to, že index tržby je roven 1. Z toho plyne, že index prodaného množství prášku se rovná převrácené hodnotě cenového indexu, který je 1,05. V červnu se tedy prodalo prášku Nela o 4,8 % méně než v květnu.

□

Příklad 8.6

V tabulce jsou údaje o cenách, prodaném množství a tržbách z prodeje určitého druhu čaje v letech 2009 – 2012. Spočtěte všechny individuální indexy (základní období je rok 2009).

Rok	Cena (Kč/ks)	Množství (tis. ks)	Tržba (tis. Kč)
2009	40	20	800
2010	42	18	756
2011	43	19	820
2012	44	17	748

Řešení:
Bude-li rok 2010 základním obdobím a rok 2009 obdobím běžným, je individuální cenový index

$$Ip = \frac{42}{40} = 1,05\,.$$

Cena se zvýšila o 5 %. Obdobně individuální index množství je

$$Iq = \frac{18}{20} = 0,9\,.$$

Prodané množství pokleslo o 10 %. Individuální index tržby je

$$IQ = \frac{756}{800} = 0,945\,,$$

tržba klesla o 5,5 %.
Budeme-li i dále používat rok 2009 jako základní období a ostatní roky jako období běžná, můžeme vypočtené jednoduché indexy zapsat do tabulky jako bazické indexy ceny, množství a tržby se základem v roce 2009 (viz následující tabulka).

Rok	*Ip*	*Iq*	*IQ*
2009	1,00	1,00	1,000
2010	1,05	0,90	0,945
2011	1,08	0,95	1,026
2012	1,10	0,85	0,935

Povšimněme si, že mezi individuálním indexem tržby, ceny a množství musí opět platit

$$IQ = \frac{Q_1}{Q_0} = \frac{p_1 \cdot q_1}{p_0 \cdot q_0} = \frac{p_1}{p_0} \cdot \frac{q_1}{q_0} = Ip \cdot Iq\,,$$

v procentech

$$IQ \cdot 100 = \frac{Q_1}{Q_2} \cdot 100 = \frac{p_1 q_1}{p_0 q_0} \cdot 100 = \frac{p_1}{p_0} \cdot 100 \cdot \frac{q_1}{q_0} \cdot 100/100 = \frac{Ip \cdot Iq}{100}\,.$$

Například pro rok 2012:

$$1{,}1 \cdot 0{,}85 = 0{,}935, \text{ resp. v procentech } 110 \cdot 85 / 100 = 93{,}5\ \%.$$

□

8.2.2 Individuální indexy složené

Příklad 8.7

V následující tabulce je uvedena průměrná měsíční mzda zaměstnance, počet zaměstnanců a měsíční mzdový fond ve čtyřech pobočkách určité firmy v lednu roku 2011 (základní období) a v lednu roku 2010 (běžné období). Jak se změnily tyto ukazatele v jednotlivých pobočkách a jak za celou firmu v roce 2011 oproti roku 2010?

Pobočka	Prům. měs. mzda (Kč)		Počet zaměstnanců		Měs. mzd. fond (Kč)	
	p_0	p_1	q_0	q_1	$Q_0 = p_0 \cdot q_0$	$Q_1 = p_1 \cdot q_1$
1	18 200	18 800	20	25	364 000	470 000
2	17 800	18 200	30	25	534 000	455 000
3	18 000	18 400	40	50	720 000	920 000
4	17 600	18 000	40	40	704 000	720 000
Součet	×	×	130	140	2 322 000	2 565 000

Řešení:

Změny za jednotlivé pobočky jsou shrnuty v následující tabulce.

Pobočka	Prům. měs. mzda (Kč)		Počet zaměstnanců		Měs. mzd. fond (Kč)	
	Ip	Δp	Iq	Δq	IQ	ΔQ
1	1,033	600	1,250	5	1,291	106 000
2	1,022	400	0,833	-5	0,852	-79 000
3	1,022	400	1,250	10	1,278	200 000
4	1,023	400	1,000	0	1,023	16 000
Firma	1,026	460	1,077	10	1,05	243 000

Například v první pobočce je index a absolutní rozdíl průměrné měsíční mzdy

$$\frac{p_1}{p_0} = \frac{18\,800}{18\,200} = 1{,}033, \quad p_1 - p_0 = 18\,800 - 18\,200 = 600\,;$$

index a absolutní rozdíl počtu zaměstnanců je

$$\frac{q_1}{q_0} = \frac{25}{20} = 1{,}250\,,\quad q_1 - q_0 = 25 - 20 = 5\,;$$

a konečně index a absolutní rozdíl pro měsíční mzdový fond je

$$\frac{Q_1}{Q_0} = \frac{470\,000}{364\,000} = 1{,}291\,,\quad Q_1 - Q_0 = 470\,000 - 364\,000 = 106\,000\,.$$

Pro celou firmu nejprve shrneme stejnorodé veličiny (počet zaměstnanců, měsíční mzdový fond, průměrná měsíční mzda) z prostorového hlediska (za různé pobočky). Stejnorodé extenzitní veličiny shrnujeme součtem ($\Sigma Q, \Sigma q$); indexy (resp. absolutní rozdíly) extenzitních veličin jsou dány vztahy:

$$I(\sum Q) = \frac{\sum Q_1}{\sum Q_0} = \frac{\sum IQ \cdot Q_0}{\sum Q_0} = \frac{\sum Q_1}{\sum \frac{Q_1}{IQ}}\,,$$

$$\Delta(\sum Q) = \sum Q_1 - \sum Q_2\,,$$

tedy pro mzdový fond

$$I(\sum Q) = \frac{\sum Q_1}{\sum Q_0} = \frac{2\,565\,000}{2\,322\,000} = 1{,}105\,.$$

$$\Delta(\sum Q) = 2\,565\,000 - 2\,322\,000 = 243\,000\,.$$

Podobně

$$I(\sum q) = \frac{\sum q_1}{\sum q_0} = \frac{\sum Iq \cdot q_0}{\sum q_0} = \frac{\sum q_1}{\sum \frac{q_1}{Iq}}\,,$$

$$\Delta(\sum q) = \sum q_1 - \sum q_0\,,$$

tedy pro počet zaměstnanců

$$I(\sum q) = \frac{\sum q_1}{\sum q_0} = \frac{140}{130} = 1{,}077\,.$$

$$\Delta(\sum q) = 140 - 130 = 10\,.$$

Počet zaměstnanců vzrostl proti lednu 2011 o 7,7 %, tj. o 10 lidí, a měsíční mzdový fond se zvýšil o 10,5 %, tj. o 243 000 Kč.
Stejnorodé intenzitní veličiny shrnujeme průměrem; v závislosti na tom, které veličiny máme k dispozici, ho určíme jako

$$\bar{p} = \frac{\sum Q}{\sum q} = \frac{\sum p.q}{\sum q} = \frac{\sum Q}{\sum \frac{Q}{p}}\,.$$

Index průměrné intenzitní veličiny (tzv. index proměnlivého složení) má potom tvar

$$I\bar{p} = \frac{\bar{p}_1}{\bar{p}_0} = \frac{\dfrac{\sum Q_1}{\sum q_1}}{\dfrac{\sum Q_0}{\sum q_0}} = \frac{\dfrac{\sum p_1 q_1}{\sum q_1}}{\dfrac{\sum p_0 q_0}{\sum q_0}} = \frac{\dfrac{\sum Q_1}{\sum \frac{Q_1}{p_1}}}{\dfrac{\sum Q_0}{\sum \frac{Q_0}{p_0}}}\,.$$

V příkladu ho můžeme určit takto:

$$I\bar{p} = \frac{\bar{p}_1}{\bar{p}_0} = \frac{\dfrac{\sum Q_1}{\sum q_1}}{\dfrac{\sum Q_0}{\sum q_0}} = \frac{\dfrac{2\,565\,000}{140}}{\dfrac{2\,322\,000}{130}} = \frac{18\,321{,}4}{17\,861{,}5} = 1{,}026\,.$$

Pro odpovídající diferenci platí

$$\Delta\bar{p} = \bar{p}_1 - \bar{p}_0 = \frac{\sum Q_1}{\sum q_1} - \frac{\sum Q_0}{\sum q_1} = \frac{\sum p_1 q_1}{\sum q_1} - \frac{\sum p_0 q_0}{\sum q_0} = \frac{\sum Q_1}{\sum \frac{Q_1}{p_1}} - \frac{\sum Q_0}{\sum \frac{Q_0}{p_0}}\,,$$

V příkladu tedy

$$\Delta\bar{p} = \bar{p}_1 - \bar{p}_0 = 18\,321{,}4 - 17\,861{,}5 = 459{,}9\,.$$

Průměrná měsíční mzda zaměstnance firmy vzrostla ve srovnání s lednem roku 2010 o 2,6 %, neboli o cca 460 Kč.

▫

8.3 Souhrnné indexy

8.3.1 Souhrnné indexy úrovně

Souhrnné indexy pro intenzitní veličiny se označují jako indexy úrovně. Nejčastěji se používají pro srovnávání cen, jedná se pak o souhrnné indexy cenové.

Příklad 8.8
Ceny a prodané množství pěti druhů zboží v červnu (základní období) a prosinci (běžné období) roku 2011 jsou uvedeny v následující tabulce.

	Cena		Množství		Pomocné výpočty			
Zboží	p_0	p_1	q_0	q_1	p_0q_0	p_1q_1	p_0q_1	p_1q_0
A	8	10	30	20	240	200	160	300
B	4	6	50	40	200	240	160	300
C	5	8	50	30	250	240	150	400
D	7	7	30	20	210	140	140	210
E	9	8	10	20	90	160	180	80
Součet	×	×	×	×	990	980	790	1 290

Určete pomocí souhrnných cenových indexů, jak se změnily ceny všeho uvedeného zboží v prosinci oproti červnu.

Řešení:
Při konstrukci souhrnných indexů se snažíme vyjádřit pomocí jednoho čísla změnu veličiny, jejíž složky jsou různého typu a jsou vyjádřeny v různých měrných jednotkách; extenzitní veličiny nelze tedy sčítat a v důsledku toho není možné ani průměrovat hodnoty odvozených intenzitních veličin.
(Poznámka: Údaje v tabulkách ilustrujících výpočty jsou dále uváděny na tři desetinná místa; výpočty jsou však prováděny – především v případě použití logaritmů – přesněji, aby nebyly zaokrouhlením deformovány příslušné vztahy mezi indexy či diferencemi; výsledky jsou pak zaokrouhleny.)

Paascheho cenový index

$$Ip^{(P)} = \frac{\sum p_1 q_1}{\sum p_0 q_1} = \frac{\sum p_1 q_1}{\sum p_1 q_1 / \left(\dfrac{p_1}{p_0} \right)}$$

vyjadřuje relativní změnu cen při stálém objemu prodeje odpovídajícím běžnému období. V příkladu

$$Ip^{(P)} = \frac{\sum p_1 q_1}{\sum p_0 q_1} = \frac{980}{790} = 1{,}241 .$$

Laspeyresův cenový index

$$Ip^{(L)} = \frac{\sum p_1 q_0}{\sum p_0 q_0} = \frac{\sum \dfrac{p_1}{p_0} p_0 q_0}{\sum p_0 q_0} ,$$

určuje relativní změnu cen při stálém objemu prodeje odpovídajícím základnímu období. V příkladu je

$$Ip^{(L)} = \frac{\sum p_1 q_0}{\sum p_0 q_0} = \frac{1\,290}{990} = 1{,}303 .$$

Podle Paascheho cenového indexu, tj. vezmeme-li v úvahu prodané množství na úrovni prosince, došlo v prosinci oproti červnu k růstu cen o 24,1 %.
Avšak podle Laspeyresova cenového indexu, tj. vezmeme-li v úvahu prodané množství na úrovni června, vzrostly ceny v prosinci oproti červnu o 30,3 %.

Fisherův cenový index

$$Ip^{(F)} = \sqrt{Ip^{(P)} \cdot Ip^{(L)}}$$

je geometrickým průměrem Paascheho a Laspeyresova cenového indexu. V příkladu vychází

$$Ip^{(F)} = \sqrt{1{,}241 \cdot 1{,}303} = 1{,}272 .$$

Montgomeryho cenový index má tvar

$$Ip^{(M)} = \left(\frac{\sum Q_1}{\sum Q_0} \right)^{\frac{\Delta(\sum Q)_p}{\Delta(\sum Q)}},$$

kde

$$\Delta(\sum Q) = \sum Q_1 - \sum Q_0,$$

$$\Delta(\sum Q)_p = \sum \frac{\ln \frac{p_1}{p_0}}{\ln \frac{Q_1}{Q_0}} (Q_1 - Q_0).$$

Pro jeho stanovení v příkladu musíme provést ještě další pomocné výpočty – viz následující tabulka:

Zboží	$\frac{p_1}{p_0} = Ip$	$\frac{Q_1}{Q_0} = IQ$	$\Delta Q = Q_1 - Q_0$	$\frac{\ln Ip}{\ln IQ} \Delta Q$
A	1,250	0,833	− 40	48,956
B	1,500	1,200	40	88,956
C	1,600	0,960	−10	115,135
D	1,000	0,667	−70	0
E	0,889	1,778	70	-14,330
Součet	×	×	−10	238,717

Z tabulky plyne, že

$$\Delta(\sum Q) = -10,$$

$$\Delta(\sum Q)_p = 238{,}717$$

a Montgomeryho cenový index

$$Ip^{(M)} = \left(\frac{\sum Q_1}{\sum Q_0} \right)^{\frac{\Delta(\sum Q)_p}{\Delta(\sum Q)}} = \left(\frac{980}{990} \right)^{\frac{238{,}717}{-10}} = 1{,}274.$$

Fisherův index představuje tedy nárůst cen uvedeného zboží o 27,2 % a Montgomeryho index růst cen o 27,4 % (rozdíl mezi Fisherovým a Montgomeryho indexem je obvykle relativně malý).

▫

8.3.2 Souhrnné indexy množství

Souhrnné indexy množství se označují také jako indexy objemové. Nejčastěji se používají při porovnávání prodávaného množství nestejnorodých výrobků.

Příklad 8.9 – pokračování
Určete pomocí souhrnných objemových indexů, jak se změnilo množství prodaného zboží v prosinci (běžné období) oproti červnu (základní období).

	Cena		Množství		Pomocné výpočty			
Zboží	p_0	p_1	q_0	q_1	p_0q_0	p_1q_1	p_0q_1	p_1q_0
A	8	10	30	20	240	200	160	300
B	4	6	50	40	200	240	160	300
C	5	8	50	30	250	240	150	400
D	7	7	30	20	210	140	140	210
E	9	8	10	20	90	160	180	80
Součet	×	×	×	×	990	980	790	1 290

Řešení:
Paascheho objemový index

$$Iq^{(P)} = \frac{\sum q_1 p_1}{\sum q_0 p_1} = \frac{\sum q_1 p_1}{\sum q_1 p_1 \Big/ \left(\dfrac{q_1}{q_0}\right)}$$

vyjadřuje relativní změnu objemu prodeje při cenové hladině odpovídající běžnému období. V příkladu

$$Iq^{(P)} = \frac{\sum q_1 p_1}{\sum q_0 p_1} = \frac{980}{1\,290} = 0,760\,.$$

Laspeyresův objemový index

$$Iq^{(L)} = \frac{\sum q_1 p_0}{\sum q_0 p_0} = \frac{\sum \frac{q_1}{q_0} q_0 p_0}{\sum q_0 p_0}$$

představuje relativní změnu objemu prodeje při cenové hladině odpovídající základnímu období. V příkladu

$$Iq^{(L)} = \frac{\sum q_1 p_0}{\sum q_0 p_0} = \frac{790}{990} = 0{,}800\,.$$

Podle Paascheho objemového indexu, tj. bereme-li v úvahu prosincové ceny, kleslo množství prodaného zboží v prosinci proti červnu o 24 %.
Avšak podle Laspeyresova objemového indexu, tj. bereme-li v úvahu červnové ceny, kleslo množství prodaného zboží v prosinci proti červnu o 20 %.

Fisherův objemový index

$$Iq^{(F)} = \sqrt{Iq^{(P)} \cdot Iq^{(L)}}$$

je geometrickým průměrem Paascheho a Laspeyresova objemového indexu; v příkladu

$$Iq^{(F)} = \sqrt{0{,}760 \cdot 0{,}800} = 0{,}780\,.$$

Montgomeryho objemový index má tvar

$$Iq^{(M)} = \left(\frac{\sum Q_1}{\sum Q_0} \right)^{\frac{\Delta(\sum Q)_q}{\Delta(\sum Q)}},$$

kde

$$\Delta(\sum Q) = \sum Q_1 - \sum Q_0\,,$$

$$\Delta(\sum Q)_q = \sum \frac{\ln \frac{q_1}{q_0}}{\ln \frac{Q_1}{Q_0}} (Q_1 - Q_0),$$

Potřebné pomocné výpočty obsahuje následující tabulka. Dostáváme v ní

$$\Delta(\sum Q) = -10\,,$$

$$\Delta(\sum Q)_q = -248{,}717$$

a Montgomeryho objemový index je tedy

$$Iq^{(M)} = \left(\frac{980}{990}\right)^{\frac{-248{,}717}{-10}} = 0{,}777\,.$$

Fisherův index představuje pokles množství prodaného zboží o 22 % a Montgomeryho index pokles o 22,3 %.

Zboží	$\dfrac{q_1}{q_0} = Iq$	$\dfrac{Q_1}{Q_0} = IQ$	$Q_1 - Q_0 = \Delta Q$	$\dfrac{\ln Iq}{\ln IQ}\Delta Q$
A	0,667	0,833	−40	−88,956
B	0,800	1,200	40	−48,956
C	0,600	0,960	−10	−125,135
D	0,667	0,667	−70	−70,000
E	2,000	1,778	70	84,330
Součet	×	×	−10	−248,717

Poznámka: Součet posledních sloupců v obou tabulkách s pomocnými výpočty pro stanovení Montgomeryho cenového a objemového indexu musí být roven ΔQ; tak například pro zboží A: 48,956 + (−88,956) = −40, pro zboží B: 88,956 + (− 48, 956) = 40 apod.

◘

Příklad 8.10

Na základě údajů v následující tabulce o tržbách z prodeje čtyř druhů mléčných výrobků v roce 2011 a 2012 (v tis. Kč) a o změnách cen v roce 2012 oproti roku 2011 určete Fisherův souhrnný cenový a Fisherův souhrnný objemový index.

Výrobek	$I(p)=\dfrac{p_1}{p_0}$	$Q_0 = p_0q_0$	$Q_1 = p_1q_1$
1.	1,20	600	576,0
2.	1,12	400	425,6
3.	1,08	500	648,0
4.	1,02	500	479,4
Součet	×	2 000	2 129

Řešení:
Nejprve určeme souhrnný cenový index Laspeyresův v průměrovém tvaru (pomocné výpočty jsou v následující tabulce):

Zboží	$\dfrac{p_1}{p_0}\cdot p_0q_0$	$\dfrac{p_1q_1}{p_1/p_0}$
U	720	480
V	448	380
X	540	600
Y	510	470
Součet	2 218	1 930

$$Ip^{(L)} = \frac{\sum \frac{p_1}{p_0} p_0q_0}{\sum p_0q_0} = \frac{2\,218}{2\,000} = 1{,}109\,.$$

Dále souhrnný cenový index Paascheho v průměrovém tvaru

$$Ip^{(P)} = \frac{\sum p_1q_1}{\sum \frac{p_1q_1}{p_1/p_0}} = \frac{2\,129}{1\,930} = 1{,}103\,.$$

Konečně Fisherův index bude $\sqrt{1{,}109 \cdot 1{,}103} = 1{,}106$ a vyjadřuje souhrnně nárůst cen o 10,6 %.

Analogicky pro souhrnné indexy objemové. Laspeyresův index

$$I_q^{(L)} = \frac{\sum p_0 q_1}{\sum p_0 q_0} = \frac{1\,930}{2\,000} = 0{,}965\,.$$

Paascheho index

$$I_q^{(P)} = \frac{\sum p_1 q_1}{\sum p_1 q_0} = \frac{2\,129}{2\,218} = 0{,}960\,.$$

Fisherův index potom bude $\sqrt{0{,}965 \cdot 0{,}960} = 0{,}962$ a vyjadřuje pokles prodaného množství o 3,8 %.

◘

8.4 Analýza indexů a absolutních rozdílů

8.4.1 Rozklad indexu a absolutního rozdílu průměrného intenzitního ukazatele

Příklad 8.11

a) Na základě údajů v tabulce určete, jak se změnila průměrná cena, počet prodaných kusů a celková tržba z prodeje oplatek Artemis v květnu (běžné období) oproti lednu (základní období) ve čtyřech sledovaných prodejnách.

b) Určete vliv změny cen v jednotlivých prodejnách a vliv změny struktury prodeje na změnu průměrné ceny.

Prodejna	Cena (Kč)		Prodáno (ks)		Tržba (Kč)	
	p_0	p_1	q_0	q_1	Q_0	Q_1
1	16	19	200	250	3 200	4 750
2	18	20	150	120	2 700	2 400
3	20	22	80	50	1 600	1 100
4	20	20	70	80	1 400	1 600
Součet	×	×	500	500	8 900	9 850

Řešení:

a) Výpočet indexů a absolutních rozdílů

$$I(\sum q) = \frac{\sum q_1}{\sum q_0} = \frac{500}{500} = 1,$$

$$\Delta(\sum q) = \sum q_1 - \sum q_0 = 500 - 500 = 0,$$

$$I(\sum Q) = \frac{\sum Q_1}{\sum Q_0} = \frac{9\,850}{8\,900} = 1{,}107,$$

$$\Delta(\sum Q) = \sum Q_1 - \sum Q_0 = 9\,850 - 8\,900 = 950,$$

$$I\overline{p} = \frac{\overline{p}_1}{\overline{p}_0} = \frac{\dfrac{\sum Q_1}{\sum q_1}}{\dfrac{\sum Q_0}{\sum q_0}} = \frac{\dfrac{9\,850}{500}}{\dfrac{8\,900}{500}} = \frac{19{,}70}{17{,}80} = 1{,}107,$$

$$\Delta\overline{p} = \overline{p}_1 - \overline{p}_0 = 19{,}70 - 17{,}80 = 1{,}90.$$

Celkový objem prodeje se v květnu oproti lednu nezměnil (vždy 500 kusů); celková tržba vzrostla o 10,7 %, tj. o 950 Kč. Průměrná cena tak vzrostla rovněž o 10,7 %, a sice o 1,90 Kč za kus.

b) Vliv změny samotných cen a vliv změny struktury prodeje určíme na základě rozkladu indexu proměnlivého složení. Použít můžeme metodu postupných změn, metodu rozkladu se zbytkem, případně logaritmickou metodu rozkladu.

Metoda postupných změn

Index proměnlivého složení lze rozložit s užitím této metody při zvoleném pořadí změn analytických veličin p, s (kdy $s = q / \Sigma q$) na součin indexu stálého složení a indexu struktury:

$$I\overline{p} = \frac{\dfrac{\sum p_1 q_0}{\sum q_0}}{\dfrac{\sum p_0 q_0}{\sum q_0}} \cdot \frac{\dfrac{\sum p_1 q_1}{\sum q_1}}{\dfrac{\sum p_1 q_0}{\sum q_0}} = \frac{\sum p_1 s_0}{\sum p_0 s_0} \cdot \frac{\sum p_1 s_1}{\sum p_1 s_0}.$$

Přitom struktura prodeje v lednu, resp. v květnu je

$$\frac{q_0}{\sum q_0} = s_0 \quad \text{a} \quad \frac{q_1}{\sum q_1} = s_1 .$$

Prodejna	p_1q_0	p_0q_1	$s_0 = \frac{q_0}{\sum q_0}$	$s_1 = \frac{q_1}{\sum q_1}$	p_1s_0	p_0s_1
1	3 800	4 000	0,40	0,50	7,60	8,00
2	3 000	2 160	0,30	0,24	6,00	4,32
3	1 760	1 000	0,16	0,10	3,52	2,00
4	1 400	1 600	0,14	0,16	2,80	3,20
Součet	9 960	8 760	1,00	1,00	19,92	17,52

$$I_{\bar{p}} = \frac{\frac{9\,960}{500}}{\frac{8\,900}{500}} \cdot \frac{\frac{9\,850}{500}}{\frac{9\,960}{500}} = \frac{19{,}92}{17{,}80} \cdot \frac{19{,}70}{19{,}92} = 1{,}119 \cdot 0{,}989 = 1{,}107 .$$

Vlivem změny cen v jednotlivých prodejnách (při zachování lednové struktury prodeje) tedy došlo ke zvýšení průměrné ceny o 11,9 %, změna struktury prodeje (při květnové cenové hladině) vyvolala pokles průměrné ceny o 1,1 %.
Rozklad odpovídajícího absolutního rozdílu je vyjádřen vztahem

$$\Delta\bar{p} = \left(\frac{\sum p_1q_0}{\sum q_0} - \frac{\sum p_0q_0}{\sum q_0} \right) + \left(\frac{\sum p_1q_1}{\sum q_1} - \frac{\sum p_1q_0}{\sum q_0} \right) =$$

$$= (\sum p_1s_0 - \sum p_0s_0) + (\sum p_1s_1 - \sum p_1s_0) .$$

$$\Delta\bar{p} = (19{,}92 - 17{,}80) + (19{,}70 - 19{,}92) = 2{,}12 + (-0{,}22) = 1{,}90 .$$

Vlivem změny cen v jednotlivých prodejnách (při zachování lednové struktury prodeje) došlo ke zvýšení průměrné ceny jednoho balení o 2,12 Kč; změna struktury prodeje (při květnové cenové hladině) však vyvolala pokles průměrné ceny o 0,22 Kč.
S užitím metody postupných změn při zvoleném pořadí změn analytických veličin *s, p* lze index proměnlivého složení rozložit na součin indexu struktury a indexu stálého složení

$$\bar{Ip}=\frac{\dfrac{\sum p_0q_1}{\sum q_1}}{\dfrac{\sum p_0q_0}{\sum q_0}}\cdot\frac{\dfrac{\sum p_1q_1}{\sum q_1}}{\dfrac{\sum p_0q_1}{\sum q_1}}=\frac{\sum p_0s_1}{\sum p_0s_0}\cdot\frac{\sum p_1s_1}{\sum p_0s_1},$$

$$\bar{Ip}=\frac{\dfrac{8\,760}{500}}{\dfrac{8\,900}{500}}\cdot\frac{\dfrac{9\,850}{500}}{\dfrac{8\,760}{500}}=\frac{17{,}52}{17{,}80}\cdot\frac{19{,}70}{17{,}52}=0{,}984\cdot 1{,}124=1{,}107\,.$$

Změna struktury prodeje zboží při lednových cenách snižuje v tomto případě průměrnou cenu o 1,6 %, důsledkem změn cen v jednotlivých prodejnách je zvýšení průměrné ceny o 12,4 %.
Rozklad odpovídajícího absolutního rozdílu lze zapsat jako

$$\Delta\bar{p}=\left(\frac{\sum p_0q_1}{\sum q_1}-\frac{\sum p_0q_0}{\sum q_0}\right)+\left(\frac{\sum p_1q_1}{\sum q_1}-\frac{\sum p_0q_1}{\sum q_1}\right)=$$

$$=\left(\sum p_0s_1-\sum p_0s_0\right)+\left(\sum p_1s_1-\sum p_0s_1\right),$$

$$\Delta\bar{p}=(17{,}52-17{,}80)+(19{,}70-17{,}52)=-0{,}28+2{,}18=1{,}90\,.$$

Změna struktury prodeje při lednových cenách snižuje průměrnou cenu o 0,28 Kč, důsledkem změn cen v jednotlivých prodejnách je však zvýšení průměrné ceny o 2,18 Kč.

Rozklad se zbytkem

Chceme-li eliminovat nejednoznačnost rozkladu při použití metody postupných změn, lze index proměnlivého složení $\bar{Ip}$ rozložit na součin indexu stálého složení, indexu struktury a zbytkového indexu I_Z,

$$\bar{Ip}=\frac{\dfrac{\sum p_1q_0}{\sum q_0}}{\dfrac{\sum p_0q_0}{\sum q_0}}\cdot\frac{\dfrac{\sum p_0q_1}{\sum q_1}}{\dfrac{\sum p_0q_0}{\sum q_0}}\cdot I_Z=\frac{\sum p_1s_0}{\sum p_0s_0}\cdot\frac{\sum p_0s_1}{\sum p_0s_0}\cdot I_Z\,.$$

Zbytek ovšem představuje nevysvětlenou část změny průměrné ceny. V tomto případě tedy dostáváme

$$I\overline{p} = \frac{19,92}{17,80} \cdot \frac{17,52}{17,80} \cdot I_Z = 1,119 \cdot 0,984 \cdot 1,005 .$$

Odpovídající rozklad absolutního rozdílu je potom

$$\Delta\overline{p} = \left(\sum p_1 s_0 - \sum p_0 s_0\right) + \left(\sum p_0 s_1 - \sum p_0 s_0\right) + \Delta_Z ,$$

zde zbytek je Δ_Z. Tedy

$$\Delta\overline{p} = (19,92 - 17,80) + (17,52 - 17,80) + \Delta_Z = 2,12 + (-0,28) + 0,06 .$$

Vlivem změny cen průměrná cena vzrostla o 11,9 %, tj. o 2,12 Kč při lednové struktuře prodeje. Změna struktury prodeje při lednových cenách znamená pokles průměrné ceny o 1,6 % (0,28 Kč). Nevysvětlen zůstává v tomto případě pouze nepatrný nárůst průměrné ceny o 0,06 Kč (0,5 %).

Logaritmická metoda rozkladu

Touto metodou rozložíme index proměnlivého složení na součin indexu stálého složení a struktury

$$I\overline{p} = I\overline{p}^{\frac{\Delta\overline{p}_s}{\Delta\overline{p}}} \cdot I\overline{p}^{\frac{\Delta\overline{p}_p}{\Delta\overline{p}}} ,$$

kde

$$\Delta\overline{p} = \overline{p}_1 - \overline{p}_0 = \Delta\overline{p}_s + \Delta\overline{p}_p ,$$

$$\Delta\overline{p}_s = \sum \frac{\ln \frac{s_1}{s_0}}{\ln \frac{p_1 s_1}{p_0 s_0}} (p_1 s_1 - p_0 s_0) \quad \text{a} \quad \Delta\overline{p}_p = \sum \frac{\ln \frac{p_1}{p_0}}{\ln \frac{p_1 s_1}{p_0 s_0}} (p_1 s_1 - p_0 s_0) .$$

První sčítanec zde představuje absolutní změnu průměrné ceny vysvětlenou vlivem změny struktury prodeje a druhý sčítanec absolutní změnu průměrné ceny vysvětlenou vlivem změn cen v jednotlivých prodejnách.

Víme již, že

$$\Delta\overline{p} = 1,90 ,$$

$$I\bar{p} = 1{,}107\,.$$

Výpočty potřebné pro analýzu jsou shrnuty v následujících tabulkách

Prodejna	p_0	p_1	q_0	q_1	$s_0 = \frac{q_0}{\sum q_0}$	$s_1 = \frac{q_1}{\sum q_1}$
1	16	19	200	250	0,40	0,50
2	18	20	150	120	0,30	0,24
3	20	22	80	50	0,16	0,10
4	20	20	70	80	0,14	0,16
Součet	x	x	500	500	1,00	1,00

Prodejna	$\ln\frac{s_1}{s_0}$	$\ln\frac{p_1}{p_0}$	$\ln\frac{p_1 s_1}{p_0 s_0}$	$\Delta(ps) = p_1 s_1 - p_0 s_0$
1	0,223144	0,171850	0,394994	3,10
2	−0,223144	0,105361	−0,117783	−0,60
3	−0,470004	0,095310	−0,374693	−1,00
4	0,133531	0,000000	0,133531	0,40
Součet	×	×	×	1,90

Prodejna	$\frac{\ln\frac{s_1}{s_0}}{\ln\frac{p_1 s_1}{p_0 s_0}}\Delta(ps)$	$\frac{\ln\frac{p_1}{p_0}}{\ln\frac{p_1 s_1}{p_0 s_0}}\Delta(ps)$
1	1,751	1,349
2	−1,137	0,537
3	−1,254	0,254
4	0,400	0,000
Součet	−0,240	2,140

Rozklad přírůstku průměrné ceny 1,90 Kč (potřebný také pro výpočet exponentů obou analytických indexů je tedy)

$$\Delta\bar{p}_s = -0,240\,,$$

$$\Delta\bar{p}_p = 2,140\,.$$

Konečně index struktury a index stálého složení jsou při použití této metody rovny

$$I\bar{p}_s = I\bar{p}^{\frac{\Delta\bar{p}_s}{\Delta\bar{p}}} = 1,107^{\frac{-0,24}{1,90}} = 0,987\,,$$

$$I\bar{p}_p = I\bar{p}^{\frac{\Delta\bar{p}_p}{\Delta\bar{p}}} = 1,107^{\frac{2,14}{1,90}} = 1,121\,.$$

Průměrná cena se důsledkem změny struktury prodeje snížila o 1,3 %, tj. o 0,24 Kč; důsledkem změny cen v jednotlivých prodejnách se průměrná cena zvýšila o 12,1 %, tj. o 2,14 Kč.

□

8.4.2 Rozklad souhrnného hodnotového indexu a diference

Příklad 8.12
Z údajů o objemu prodeje a ceně jednoho balení různých prostředků na mytí nádobí (A, B a C) určitého výrobce v roce 2008 a v roce 2010 v tabulce určete, jak se změnily celkové tržby, a dále, jak byla tato změna ovlivněna změnami cen a jak změnami prodaného množství.

	Cena za jednotku (Kč)		Prodané množství (tis. jednotek)	
Výrobek	rok 2008	rok 2010	rok 2008	rok 2010
A	60	70	200	300
B	40	30	500	400
C	80	100	300	400

Velikost tržby zjistíme vynásobením jednotkové ceny a prodaného množství výrobku; celkové tržby jsou součtem tržeb u jednotlivých výrobků (výpočet byl proveden v úvodu této kapitoly).
Změna celkových tržeb:

$$I(\sum Q) = \frac{\sum p_1 q_1}{\sum p_0 q_0} = \frac{73\,000}{56\,000} = 1{,}304\,,$$

$$\Delta(\sum Q) = \sum p_1 q_1 - \sum p_0 q_0 = 73\,000 - 56\,000 = 17\,000\,.$$

Celkové tržby vzrostly o 17 000 tis. Kč, tj. o 30,4 %.
Jak byla celková změna tržeb ovlivněna změnami cen a jak změnami prodaného množství, zjistíme na základě rozkladu indexu (a diference) celkových tržeb.

__**Metoda postupných změn**__

a) Uvažujeme-li v rozkladu nejprve změnu cen a pak změnu prodaného množství, dostaneme

$$I(\sum Q) = \frac{\sum p_1 q_1}{\sum p_0 q_0} = \frac{\sum p_1 q_0}{\sum p_0 q_0} \cdot \frac{\sum p_1 q_1}{\sum p_1 q_0} = Ip^{(L)} \cdot Iq^{(P)}.$$

Pomocné výpočty obsahuje následující tabulka:

Výrobek	Tržba (2008) p_0q_0	Tržba (2010) p_1q_1	Pomocné výpočty p_1q_0	p_0q_1
A	12 000	21 000	14 000	18 000
B	20 000	12 000	15 000	16 000
C	24 000	40 000	30 000	32 000
Součet	56 000	73 000	59 000	66 000

Po dosazení tedy dostáváme:

$$I(\sum Q) = \frac{\sum p_1 q_0}{\sum p_0 q_0} \cdot \frac{\sum p_1 q_1}{\sum p_1 q_0} = \frac{59\,000}{56\,000} \cdot \frac{73\,000}{59\,000} = 1{,}054 \cdot 1{,}237\,.$$

Rozklad příslušného absolutního rozdílu lze vyjádřit vztahem

$$\Delta(\sum Q) = \left(\sum p_1 q_0 - \sum p_0 q_0\right) + \left(\sum p_1 q_1 - \sum p_1 q_0\right),$$

tedy

$$\Delta(\sum Q) = (59\,000 - 56\,000) + (73\,000 - 59\,000) = 3\,000 + 14\,000\,.$$

Změna cen tedy způsobila nárůst tržeb o 5,4 %, tj. o 3 000 tis. Kč, změna prodaného množství nárůst tržeb o 23,7 %, tj. o 14 000 tis. Kč.

b) Uvažujeme-li v rozkladu naopak nejprve změnu prodaného množství a teprve poté změnu cen, dostaneme

$$I(\sum Q) = \frac{\sum p_1 q_1}{\sum p_0 q_0} = \frac{\sum p_0 q_1}{\sum p_0 q_0} \cdot \frac{\sum p_1 q_1}{\sum p_0 q_1} = Iq^{(L)} \cdot Ip^{(P)},$$

neboli

$$I(\sum Q) = \frac{66000}{56000} \cdot \frac{73000}{66000} = 1{,}179 \cdot 1{,}106 .$$

Analogicky absolutní rozdíl celkových tržeb rozložíme takto:

$$\Delta(\sum Q) = \left(\sum p_0 q_1 - \sum p_0 q_0\right) + \left(\sum p_1 q_1 - \sum p_0 q_1\right),$$

a po dosazení dostáváme

$$\Delta(\sum Q) = (66\,000 - 56\,000) + (73\,000 - 66\,000) = 10\,000 + 7\,000 .$$

Změna cen způsobila nárůst tržeb o 10,6 %, tj. o 7000 tis. Kč, změna prodaného množství nárůst tržeb o 17,9 %, tj. o 10 000 tis. Kč.

Metoda rozkladu se zbytkem

Souhrnný index celkových tržeb lze rozložit podle vzorce

$$I(\sum Q) = \frac{\sum p_0 q_1}{\sum p_0 q_0} \cdot \frac{\sum p_1 q_0}{\sum p_0 q_0} \cdot I_Z = Iq^{(L)} \cdot Ip^{(L)} \cdot I_Z ,$$

kde zbytek I_Z je nevysvětlená část změny celkových tržeb.

$$I(\sum Q) = \frac{59\,000}{56\,000} \cdot \frac{66\,000}{56\,000} \cdot I_Z = 1{,}054 \cdot 1{,}179 \cdot 1{,}049 .$$

Odpovídající rozklad absolutního rozdílu lze zapsat jako

$$\Delta(\sum Q) = \left(\sum p_0 q_1 - \sum p_0 q_0\right) + \left(\sum p_1 q_0 - \sum p_0 q_0\right) + \Delta_Z ,$$

Δ_Z je zde nevysvětlená část absolutního přírůstku celkových tržeb.

$$\Delta(\sum Q) = (59\,000 - 56\,000) + (66\,000 - 56\,000) + \Delta_Z =$$
$$= 3\,000 + 10\,000 + 4\,000\,.$$

Při cenách roku 2008 by změna prodaného množství způsobila nárůst celkové tržby o 17,9 %, tj. 10 000 tis. Kč; změna cen při stejném prodaném množství jako v roce 2008 by způsobila zvýšení celkové tržby o 5,4 %, tj. o 3000 tis. Kč. Nevysvětlen zůstává v tomto případě další nárůst tržeb o 4,9 % (4 000 tis. Kč).

Logaritmická metoda

Rozklad absolutního rozdílu celkových tržeb 17 000 tis. Kč (potřebný také pro výpočet exponentů obou indexů) je

$$\Delta(\sum Q) = \Delta(\sum Q)_p + \Delta(\sum Q)_q\,,$$

kde

$$\Delta(\sum Q)_p = \sum \frac{\ln \frac{p_1}{p_0}}{\ln \frac{p_1 q_1}{p_0 q_0}} (p_1 q_1 - p_0 q_0)\,,$$

$$\Delta(\sum Q)_q = \sum \frac{\ln \frac{q_1}{q_0}}{\ln \frac{p_1 q_1}{p_0 q_0}} (p_1 q_1 - p_0 q_0)\,.$$

Vzorec pro rozklad souhrnného indexu celkových tržeb má pak tvar

$$I(\sum Q) = I(\sum Q)^{\frac{\Delta(\sum Q)_p}{\Delta(\sum Q)}} \cdot I(\sum Q)^{\frac{\Delta(\sum Q)_q}{\Delta(\sum Q)}}\,.$$

V následujících tabulkách nejprve opět provedeme potřebné dílčí výpočty:

Výrobek	$\ln\frac{p_1}{p_0}$	$\ln\frac{q_1}{q_0}$	$\ln\frac{p_1q_1}{p_0q_0}$	$\Delta Q = p_1q_1 - p_0q_0$
A	0,154151	0,405465	0,559616	9 000
B	−0,287682	−0,223144	−0,510826	−8 000
C	0,223144	0,287682	0,510826	16 000
Součet	×	×	×	17 000

Výrobek	$\frac{\ln Ip}{\ln IQ}\Delta Q$	$\frac{\ln Iq}{\ln IQ}\Delta Q$
A	2 479	6 521
B	−4 505	−3 495
C	6 989	9 011
Součet	4 963	12 037

Dostáváme tedy

$$\Delta(\sum Q)_p = 4\,963 \text{ a } \Delta(\sum Q)_q = 12\,037\,.$$

Index celkových tržeb je roven, jak jsme zjistili hned v prvních krocích řešení celého příkladu,

$$I(\sum Q) = 1{,}304\,.$$

Analytické indexy (tedy Montgomeryho souhrnný index cenový a objemový) potom jsou rovny

$$Ip^{(M)} = 1{,}304^{\frac{4\,963}{17\,000}} = 1{,}081\,,$$

resp.

$$Iq^{(M)} = 1{,}304^{\frac{12\,037}{17\,000}} = 1{,}207\,.$$

Celkové tržby vlivem změny cen vzrostly o 8,1 %, tj. o 4 963 tis. Kč; vlivem změny prodaného množství vzrostly o 20,7 %, tj. o 12 037 tis. Kč.
Poznamenejme ještě, že Fisherův cenový a Fisherův objemový index jsou v tomto příkladu

$$Ip^{(F)} = \sqrt{1,054 \cdot 1,106} = 1,080,$$

$$Iq^{(F)} = \sqrt{1,237 \cdot 1,179} = 1,208,$$

a zaregistrujme jejich praktickou shodu s indexy Montgomeryho.

◘

8.5 Bortkiewiczův rozklad

Příklad 8.13

V první části příkladu 8.8 jsme určili Laspeyresův souhrnný cenový index (1,303) a Paascheho souhrnný cenový index (1,241). Vysvětlete rozdíl v hodnotách obou indexů užitím Bortkiewiczova rozkladu.

Řešení:

Na základě tzv. Bortkiewiczova rozkladu lze podíl souhrnných cenových indexů Laspeyresova a Paascheho vyjádřit jako

$$B = \frac{Ip^{(P)}}{Ip^{(L)}} = \frac{\dfrac{\sum p_1 q_1}{\sum p_0 q_1}}{\dfrac{\sum p_1 q_0}{\sum p_0 q_0}} = 1 + v_{Ip} \cdot v_{Iq} \cdot r_{IpIq},$$

kde v_{Ip} je variační koeficient jednoduchých cenových indexů (p_1/p_0), v_{Iq} je variační koeficient jednoduchých objemových indexů (q_1/q_0) a r_{IpIq} je korelační koeficient mezi jednoduchými cenovými a objemovými indexy.
Následující tabulka obsahuje potřebné dílčí výpočty.

Zboží	$\frac{p_1}{p_0} = Ip$	$\frac{q_1}{q_0} = Iq$	Q_0	$Q_0 \cdot U^2$	$Q_0 \cdot V^2$	$Q_0 \cdot U \cdot V$
A	1,250	0,667	240	0,674	4,245	1,692
B	1,500	0,800	200	7,762	0	0
C	1,600	0,600	250	22,052	10,000	–14,850
D	1,000	0,667	210	19,280	3,715	8,463
E	0,889	2,000	90	15,426	129,600	–44,712
Součet	×	×	990	65,194	147,560	–49,407

Poznámka.: $U = Ip - Ip^{(L)}$, $V = Iq - Iq^{(L)}$

V samotném rozkladu potom

$$r_{IpIq} = \frac{\sum Q_0 \cdot U \cdot V}{\sqrt{\sum Q_0 \cdot U^2 \sum Q_0 \cdot V^2}} = \frac{-49{,}407}{\sqrt{65{,}194 \cdot 147{,}560}} = -0{,}504\,,$$

$$v_{Ip} = \frac{\sqrt{\dfrac{\sum U^2 \cdot Q_0}{\sum Q_0}}}{Ip^{(L)}} = \frac{\sqrt{\dfrac{65{,}194}{990}}}{1{,}303} = 0{,}197\,,$$

$$v_{Iq} = \frac{\sqrt{\dfrac{\sum V^2 \cdot Q_0}{\sum Q_0}}}{Iq^{(L)}} = \frac{\sqrt{\dfrac{147{,}560}{990}}}{0{,}8} = 0{,}483\,,$$

a tedy

$$B = \frac{1{,}241}{1{,}303} = 0{,}952 = 1 + 0{,}197 \cdot 0{,}483 \cdot (-0{,}504)\,.$$

Variační koeficient jednoduchých cenových indexů je 0,197 a variační koeficient jednoduchých objemových indexů je 0,483, a tudíž je variabilita prodaného množství jednotlivých druhů zboží vyšší než variabilita cen. Korelační koeficient mezi jednoduchými cenovými a objemovými indexy je záporný, to znamená, že při růstu cen dochází k poklesu množství prodaného zboží (a naopak). Paascheho souhrnný index je proto nižší než index Laspeyresův.

◘

Cvičení

1. Doplníme-li k příkladu 8.2 index porovnávající počet dokončených bytů v ČR v roce 2004 s jejich počtem v roce 2000 (tento index je 1,280), určete index porovnávající počet dokončených bytů v roce 2010 s rokem 2000.

2. Index proměnlivého složení v příkladu 8.7 rozložte metodou postupných změn na index stálého složení a index struktury.

3. Údaje o ceně tří základních tarifů a o počtu klientů v určité oblasti v okamžiku vstupu nového telefonního operátora na trh (základní období) a v současnosti (běžné období) obsahuje následující tabulka. Určete, jak se změnila průměrná cena tarifu a dále, jak byla tato změna ovlivněna samotnou změnou cen jednotlivých tarifů a jak ji ovlivnila změna zájmu klientů o jednotlivé tarify.

	Cena		Počet klientů (tis.)	
Tarif	z.o.	b.o.	z.o.	b.o.
T1	320	384	25	50
T2	400	480	35	20
T3	500	600	40	20

4. V příkladu 8.9 index celkových tržeb rozložte metodou postupných změn a metodou se zbytkem. Určete souhrnný cenový a objemový index Montgomeryho. Dále vysvětlete rozdíl mezi Laspeyresovým a Paascheovým souhrnným cenovým indexem na základě Bortkiewiczova rozkladu.

Výsledky

1. 1,446.

2. Pořadí změn p, s: index stálého složení 1,0363, index struktury 1,0024. Pořadí změn s, p: index stálého složení 1,0367, index struktury 1,0020.

3. Index proměnlivého složení je 1,079. Rozklad metodou postupných změn: v důsledku zdražení vzrostla průměrná cena tarifu o 20 % (index stálého složení je 1,200), v důsledku změny zájmu klientů o jednotlivé tarify poklesla průměrná cena o cca 10 % (index struktury je 0,899).

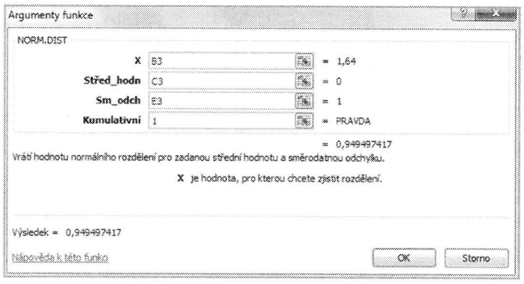
4. Index celkových tržeb 1,065;
Laspeyresův index cenový je 1,109, Paascheho index objemový 0,960;

Paascheho index cenový je 1,103, Laspeyresův index objemový je 0,965.
Montgomeryho souhrnný index cenový je 1,106 a Montgomeryho souhrnný index objemový je 0,963.
Cenové indexy se od sebe příliš neliší, je to dáno především relativně malou variabilitou ve změnách cen (variační koeficient je 0,061); korelační koeficient je záporný, závislost vývoje cen a prodaného množství je nepřímá (Paascheho index je menší než Laspeyresův).
$B = 1{,}103/1{,}109 = 1 - 0{,}061 \cdot 0{,}149 \cdot (-0{,}521)$

KAPITOLA IX

PŘÍLOHA

PRAVDĚPODOBNOSTNÍ ROZDĚLENÍ v MS EXCEL



9 I

Tato
s prav
zím
a u m
nahra
verzí
kvant
funkc
nenal
BINO
starší
verzi.
z ozn
(resp.
k funk
určitý

Dále
2010.
pravd
hodno
děné

9.1

V obl
ní, u k

Jedná
kvant

9 Pravděpodobnostní rozdělení v MS Excel

Tato příloha je v podstatě stručným manuálem, poskytujícím návod na práci s pravděpodobnostními rozděleními v MS Excel verze 10 a vyšších. Oproti předchozím verzím došlo totiž ke změnám názvu všech pravděpodobnostních funkcí a u mnoha z nich se změnil i jejich obsah. Navíc přibyly některé nové funkce, které nahrazují ty starší, přičemž byly tyto funkce rozšířeny (např. počítají oproti starším verzím nejen hodnoty distribuční funkce, ale i hodnoty hustoty pravděpodobnosti či kvantilů). Z důvodů zpětné kompatibility jsou v Excelu platné i dřívější syntaxe funkcí. To znamená, že tyto starší funkce sice fungují, ale již je v nabídce funkcí nenalezneme. Ukázkou může být např. funkce BINOM.DIST, která nahradila funkci BINOMDIST, přičemž stará i nová verze fungují úplně stejně. Naproti tomu např. ve starší verzi Excelu funkce TDIST počítá něco úplně jiného než funkce T.DIST v nové verzi. Došlo též ke sjednocení názvu funkcí, neboť nyní je název důsledně složen z označení rozdělení, pak následuje tečka a označení DIST pro distribuční funkci (resp. pravděpodobnostní funkci či hustotu) či INV (pro označení inverzní funkce k funkci distribuční, tedy pro kvantilovou funkci). Do označení funkcí tak byl vnesen určitý řád a jednotnost.

Dále popíšeme všechny funkce pravděpodobnostních rozdělení ve verzi MS Excel 2010. U každého rozdělení uvedeme vzorec pravděpodobnostní funkce či hustoty pravděpodobnosti, jak je definován v MS Excel. Dále popíšeme možnosti výpočtů hodnot distribuční funkce a kvantilů a syntaxi příslušných funkcí. Veškeré dále uváděné skutečnosti jsou vázané na MS Excel verze 10.

9.1 Diskrétní rozdělení

V oblasti diskrétních (nespojitých) rozdělení obsahuje MS Excel následující rozdělení, u kterých zároveň uvádíme název příslušné funkce:

Tab. 9.1: Přehled funkcí pro nespojitá rozdělení

rozdělení	pravděpodobnostní a distribuční funkce	kvantily
binomické	BINOM.DIST	BINOM.INV
negativně binomické	NEGBINOM.DIST	-
Poissonovo	POISSON.DIST	-
hypergeometrické	HYPGEOM.DIST	-

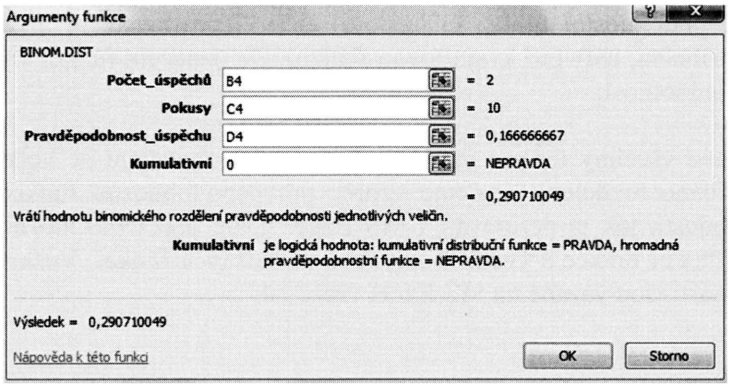
Jedná se tedy o naprosto o základní typy rozdělení, navíc ne vždy je možné spočítat kvantily. To ale není žádné neštěstí, neboť kvantily jsme schopni poměrně snadno spočítat z hodnot pravděpodobnostní funkce. Podívejme se nyní na jednotlivá rozdělení podrobněji. Je třeba ještě uvést, že distribuční funkce je v Excelu definována jako $F(x) = P(X \le x), \ \forall x \in \mathrm{R}$.

Binomické rozdělení

Náhodná veličina X má binomické rozdělení s parametry n a π, jestliže její pravděpodobnostní funkce má tvar

$$P(x) = \binom{n}{x} \pi^x (1-\pi)^{n-x}, \qquad x = 0,1,\ldots,n, \quad 0 < \pi < 1, \quad n \in \mathrm{N}.$$

V Excelu se jak pro distribuční funkci i pro pravděpodobnostní funkci používá funkce BINOM.DIST.



Její argumenty mají následující význam:

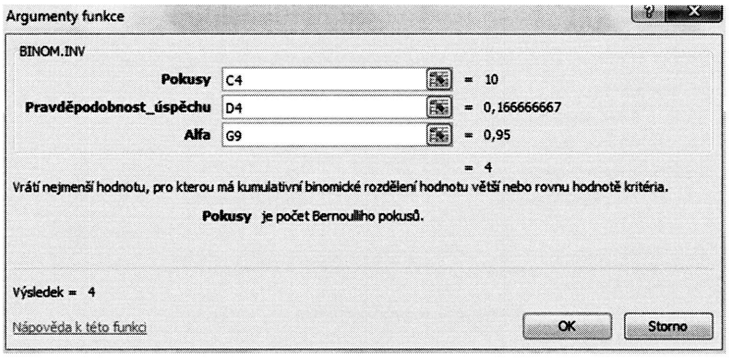
Počet_úspěchů - x (počet úspěchů). Hodnota, pro kterou počítáme $F(x)$ či $P(x)$.

Pokusy - n (počet pokusů) a parametr rozdělení.

Pravděpodobnost_úspěchu - π. Pravděpodobnost úspěchu a parametr rozdělení.

Kumulativní - NEPRAVDA pro hodnotu pravděpodobnostní funkce $P(x)$, PRAVDA pro hodnotu distribuční funkce $F(x)$.

Jako pro jediné z nespojitých rozdělení je v Excelu uvedena i funkce pro výpočet kvantilů BINOM.INV. Její argumenty mají obdobný význam, jako je tomu u funkce BINOM.DIST.



Pokusy - n (počet pokusů) a parametr rozdělení.

Pravděpodobnost_úspěchu - π. Pravděpodobnost úspěchu a parametr rozdělení.

Alfa - pravděpodobnost P pro hodnotu kvantilu x_P .

Negativně binomické rozdělení

Náhodná veličina X má negativně binomické rozdělení s parametry n a π, jestliže její pravděpodobnostní funkce má pro n celočíselné tvar

$$P(x) = \binom{n+x-1}{x} \pi^n (1-\pi)^x, \qquad x = 0,1,\ldots \quad 0 < \pi < 1, \quad n \in \mathrm{N}\,.$$

Připomeňme, že pro přirozená n můžeme náhodnou veličinu X chápat jako počet neúspěchů před n-tým úspěchem. Úmyslně uvádíme pravděpodobnostní funkci ve zjednodušeném tvaru, neboť takto je chápána v Excelu.

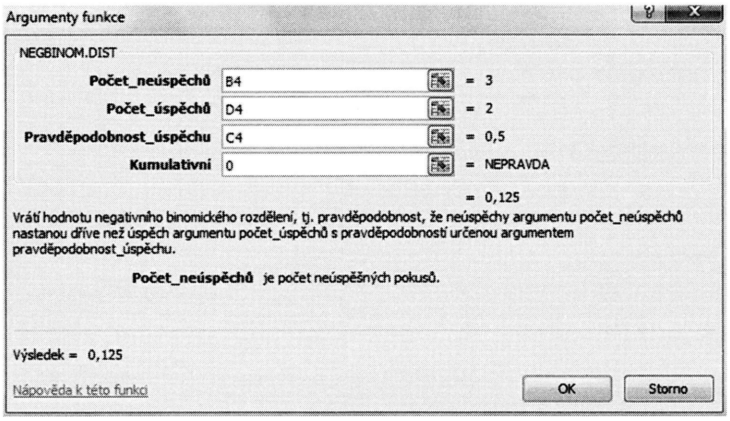
V Excelu se pro pravděpodobnostní funkci používá funkce NEGBINOM. DIST. Její argumenty mají následující význam:

Počet_neúspěchů - x (počet neúspěchů před n-tým úspěchem). Hodnota, ve které většinou počítáme $F(x)$ či $P(x)$.

Počet_úspěchů - n (počet úspěchů).

Pravděpodobnost_úspěchu - π. Pravděpodobnost úspěchu.

Kumulativní - NEPRAVDA pro hodnotu pravděpodobnostní funkce $P(x)$, PRAVDA pro hodnotu distribuční funkce $F(x)$.



Speciálním případem negativně binomického rozdělení je rozdělení geometrické. Toto rozdělení obdržíme velmi snadno, pokud v negativně binomickém rozdělení položíme $n=1$. Potom se předchozí pravděpodobnostní funkce zjednoduší do tvaru

$$P(x) = \pi(1-\pi)^x, \qquad x = 0,1,\ldots, \quad 0 < \pi < 1.$$

Náhodnou veličinu X lze potom chápat jako počet neúspěchů před prvním úspěchem.
V Excelu použijeme funkci NEGBINOMDIST, ve které položíme **Počet_úspěchů** = 1.

Poissonovo rozdělení
Náhodná veličina X má Poissonovo rozdělení s parametrem λ, jestliže její pravděpodobnostní funkce má tvar

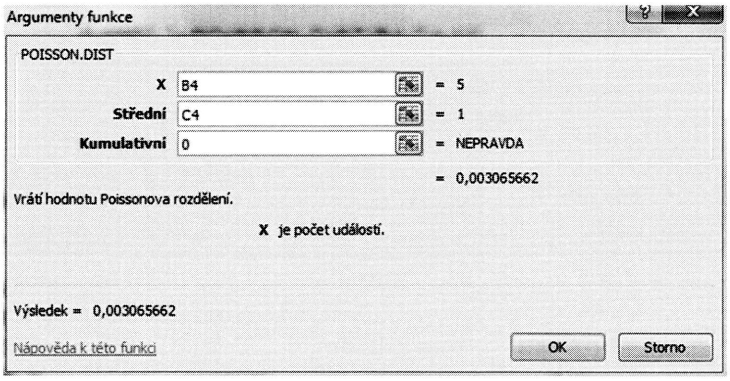
$$P(x) = \frac{\lambda^x}{x!}e^{-\lambda}, \qquad x = 0,1,\ldots, \quad \lambda > 0.$$

V Excelu se jak pro distribuční funkci i pro pravděpodobnostní funkci používá funkce POISSON.DIST Její argumenty mají následující význam:

X - x. Hodnota, ve které většinou počítáme $F(x)$ či $P(x)$.

Střední - λ. Parametr a zároveň střední hodnota rozdělení.

Kumulativní - NEPRAVDA pro hodnotu pravděpodobnostní funkce $P(x)$, PRAVDA pro hodnotu distribuční funkce $F(x)$.



Hypergeometrické rozdělení

Náhodná veličina X má hypergeometrické rozdělení s parametry N, M a n, jestliže její pravděpodobnostní funkce má tvar

$$P(x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}, \qquad x = \max(0, M-N+n), ..., \min(M,n).$$

Přitom N, M a n jsou přirozená čísla, $1 \le n < N$, $1 \le M < N$.

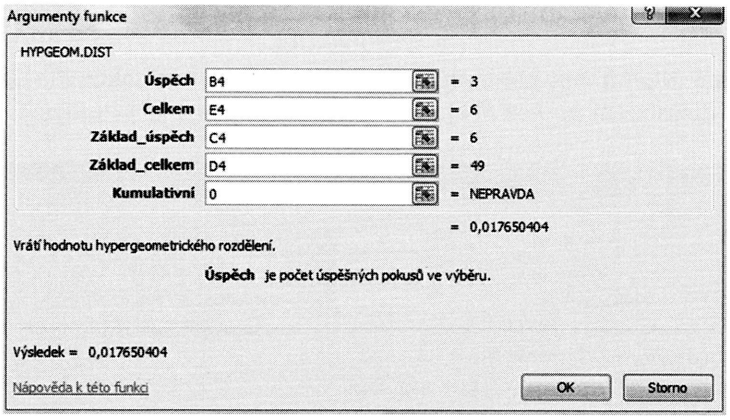
V Excelu se pro pravděpodobnostní funkci používá funkce HYPGEOM.DIST. Její argumenty mají následující význam:

Úspěch - x. Hodnota, ve které počítáme $P(x)$.

Celkem - n (rozsah výběru).

Základ_úspěch - M. Počet prvků s vlastností M.

Základ_celkem - N (rozsah základního souboru).
Kumulativní - NEPRAVDA pro hodnotu pravděpodobnostní funkce $P(x)$, PRAVDA pro hodnotu distribuční funkce $F(x)$.



Shrnutí syntaxe

Rozdělení	$F(x)$	$P(x)$	x_P
binomické	=BINOM.DIST(x;n;π;1)	=BINOM.DIST(x;n;π;0)	=BINOM.INV(n;π;P)
neg.binomické	=NEGBINOM.DIST((x;n;π;1)	=NEGBINOM.DIST((x;n;π;0)	-
Poissonovo	=POISSON.DIST(x;λ;1)	=POISSON.DIST(x;λ;0)	-
hypergeometrické	=HYPGEOM.DIST(x;n;M;N;1)	=HYPGEOM.DIST(x;n;M;N;0)	-

9.2 Spojitá rozdělení

V oblasti spojitých rozdělení obsahuje MS Excel řadu rozdělení, u kterých opět uvádíme název příslušné funkce pro výpočet distribuční funkce, hustoty a kvantilu:

Tab. 9.2: Přehled funkcí pro spojitá rozdělení

rozdělení	distribuční funkce a hustota	kvantily
normální	NORM.DIST	NORM.INV
normované normální	NORM.S.DIST	NORM.S.INV
logaritmicko normální	LOGNORM.DIST	LOGNORM.INV
exponenciální	EXPON.DIST	-
Weibullovo	WEIBULL.DIST	-
Studentovo (t)	T.DIST	T.INV
Fischer-Schnedecorovo (F)	F.DIST	F.INV
chí-kvadrát	CHISQ.DIST	CHISQ.INV
Beta	BETA.DIST	BETA.INV
Gama	GAMMA.DIST	GAMMA.INV

Další funkce jsou v následující tabulce.

Tab.9.3: Přehled funkcí pro spojitá rozdělení

rozdělení	1-$F(x)$	$P(\lvert X\rvert > x)$	$x_{1-P/2}$	x_{1-P}
Studentovo (t)	T.DIST.RT	T.DIST.2T	T.INV.2T	-
Fischer-Schnedecorovo (F)	F.DIST.RT	-	-	F.INV.RT
chí-kvadrát	CHISQ.DIST.RT	-	-	CHISQ.INV.RT

Je tedy zřejmé, že nabídka spojitých rozdělení je podstatně širší, než je tomu u rozdělení nespojitých.

Normální rozdělení

Náhodná veličina X má normální rozdělení s parametry μ a σ^2, jestliže její hustota pravděpodobnosti má tvar

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < +\infty, \quad -\infty < \mu < +\infty, \quad \sigma > 0 \; .$$

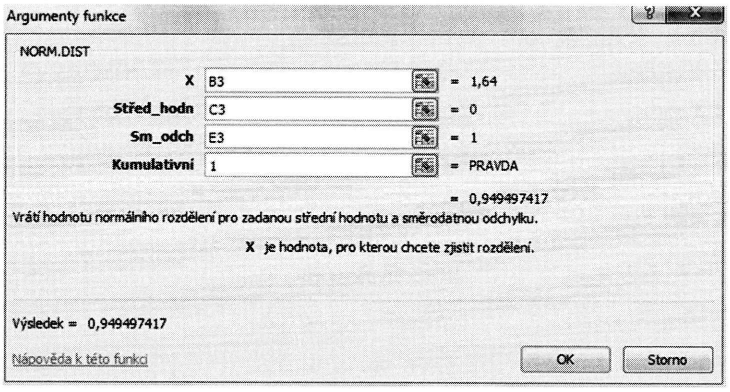
V Excelu se pro distribuční funkci a hustotu používá funkce NORM.DIST. Její argumenty mají následující význam:

X - x. Hodnota, ve které počítáme $F(x)$, resp. $f(x)$.

Střed_hodn - μ. Parametr rozdělení a zároveň střední hodnota.

Sm_odch - σ. Parametr rozdělení a zároveň odmocnina z rozptylu (tedy směrodatná odchylka).

Kumulativní - NEPRAVDA pro hodnotu hustoty $f(x)$, PRAVDA pro hodnotu distribuční funkce $F(x)$.

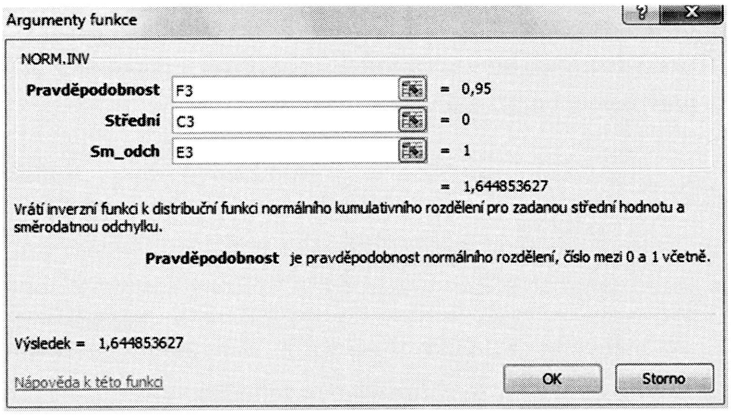


Funkce pro výpočet kvantilů normálního rozdělení má v Excelu název NORM.INV. Jedná se o kvantilovou funkci $F^{-1}(P) = x_P$, která má následující argumenty:

Pravděpodobnost - pravděpodobnost P pro hodnotu kvantilu x_P.

Střední - μ. Parametr rozdělení a zároveň střední hodnota.

Sm_odch - σ. Funkce parametru rozdělení a zároveň odmocnina z rozptylu (tedy směrodatná odchylka).



Normované normální rozdělení

Kromě normálního rozdělení s obecnými parametry μ a σ^2 nabízí Excel i normované normální rozdělení, tedy rozdělení s parametry $\mu = 0$ a $\sigma^2 = 1$.

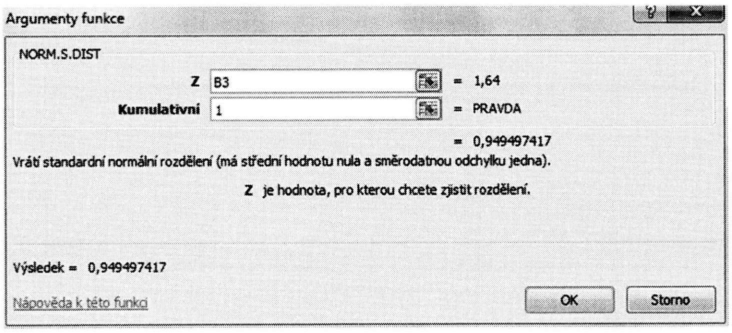
Náhodná veličina U má normální rozdělení s parametry 0 a 1, jestliže její hustota pravděpodobnosti má tvar

$$f(u) = \frac{1}{\sqrt{2\pi}} e^{\frac{-u^2}{2}}, \quad -\infty < u < +\infty \ .$$

V Excelu se pro distribuční funkci používá funkce NORM.S.DIST. Tato funkce má 2 argumenty:

Z - u. Hodnota, ve které počítáme $F(u)$.

Kumulativní - NEPRAVDA pro hodnotu hustoty $f(x)$, PRAVDA pro hodnotu distribuční funkce $F(x)$.



Logaritmicko normální rozdělení

Náhodná veličina X má logaritmicko normální rozdělení s parametry μ a σ^2, jestliže její hustota pravděpodobnosti má tvar

$$\begin{aligned} f(x) &= \frac{1}{\sigma x\sqrt{2\pi}} e^{\frac{-(\ln x-\mu)^2}{2\sigma^2}}, & x>0,\ -\infty<\mu<+\infty,\ \sigma>0, \\ &= 0, & x\le 0. \end{aligned}$$

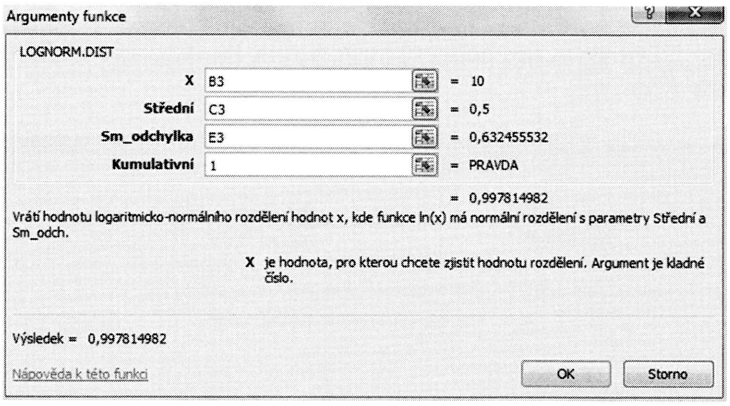
Připomeňme, že náhodná veličina $Y=\ln(X)$ má potom normální rozdělení s parametry μ a σ^2 - tedy přirozený logaritmus náhodné veličiny s logaritmicko normálním rozdělením má normální rozdělení se stejnými parametry μ a σ^2.
V Excelu se pro výpočet hodnot distribuční funkce používá funkce LOGNORM.DIST. Její argumenty mají následující význam:

X - x. Hodnota, ve které počítáme $F(x)$ či $f(x)$.

Střední - μ. Parametr rozdělení. Pozor, nejedná se o střední hodnotu X, nýbrž o střední hodnotu $\ln(X)$.

Sm_odchylka - σ. Parametr rozdělení. Opět se nejedná o směrodatnou odchylku X, nýbrž o směrodatnou odchylku hodnoty $\ln(X)$.

Kumulativní - NEPRAVDA pro hodnotu hustoty $f(x)$, PRAVDA pro hodnotu distribuční funkce $F(x)$.

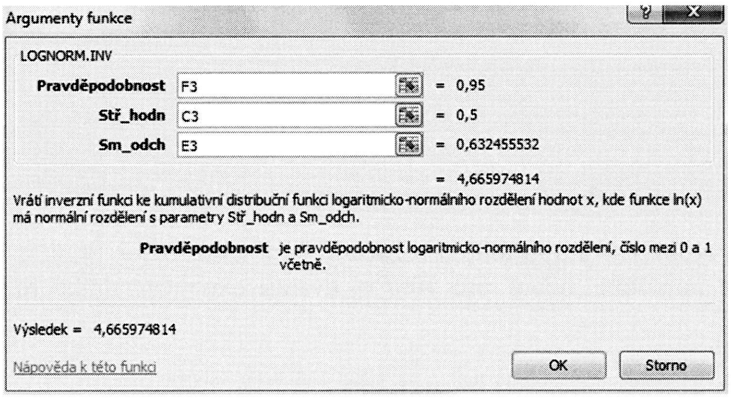


Funkce pro výpočet kvantilů logaritmicko normálního rozdělení má v Excelu název LOGNORM.INV. Jedná se o kvantilovou funkci $F^{-1}(P) = x_P$, která má následující argumenty:

Pravděpodobnost - pravděpodobnost P pro hodnotu kvantilu x_P.

Stř_hodn - μ. Parametr rozdělení a zároveň střední hodnota veličiny ln(X). Nejedná se tedy o střední hodnotu logaritmicko normálního rozdělení, jak by název parametru mohl mylně evokovovat.

Sm_odch - σ. Parametr rozdělení, odmocnina ze σ^2. Zároveň směrodatná odchylka veličiny ln(X) . Nejedná se tedy o směrodatnou odchylku logaritmicko normálního rozdělení, jak by název parametru mohl mylně evokovovat.



Exponenciální rozdělení

Náhodná veličina X má exponenciální normální rozdělení s parametrem λ, jestliže její hustota pravděpodobnosti má tvar

$$\begin{aligned} f(x) &= \lambda e^{-\lambda x}, \quad x > 0,\ \lambda > 0, \\ &= 0, \qquad\quad \mathrm{x} \le 0. \end{aligned}$$

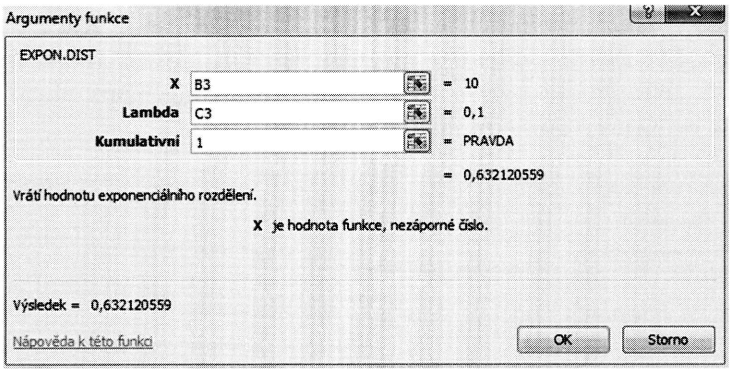
Pokud položíme $\lambda = \dfrac{1}{\delta}$, dostali bychom tvar rozdělení, v jakém je obvykle uváděn v literatuře. Nicméně v Excelu je uveden v této podobě a tento tvar je prezentován pro hodnoty $x > 0$, neuvažuje se zde tedy možné posunutí A.

V Excelu se pro výpočet hodnot distribuční funkce a hustoty používá funkce EXPON.DIST. Její argumenty mají následující význam:

X - x. Hodnota, ve které počítáme $F(x)$, resp. $f(x)$.

Lambda - λ. Parametr rozdělení.

Kumulativní - NEPRAVDA pro hodnotu hustoty $f(x)$, PRAVDA pro hodnotu distribuční funkce $F(x)$.



Funkce pro výpočet kvantilů tohoto rozdělení není k dispozici. S jejím výpočtem si však snadno poradíme, neboť pro $100P\%$ kvantil exponenciálního rozdělení platí vztah

$$x_P = \frac{-\ln(1-P)}{\lambda}, \quad 0 < P < 1$$

Weibullovo rozdělení

Náhodná veličina X má Weibullovo rozdělení s parametry δ a α, jestliže její hustota pravděpodobnosti má tvar

$$\begin{aligned} f(x) &= \frac{\alpha x^{\alpha-1}}{\beta^{\alpha}} \exp\left[-\left(\frac{x}{\beta}\right)^{\alpha}\right], && x>0,\ \beta>0,\ \alpha>0, \\ &= 0, && x \le 0. \end{aligned}$$

Speciálním případem Weibullova rozdělení je pro $\alpha = 1$ exponenciální rozdělení.

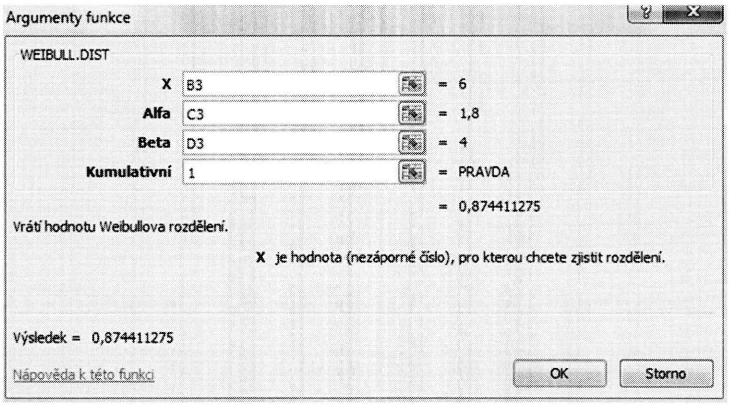
Pro výpočet hodnot distribuční funkce a hustoty se používá funkce WEIBULL.DIST. Její argumenty mají následující význam:

X - x. Hodnota, ve které počítáme $F(x)$, resp. $f(x)$.

Alfa - α. Parametr rozdělení.

Beta - β. Parametr rozdělení.

Kumulativní - NEPRAVDA pro hodnotu hustoty $f(x)$, PRAVDA pro hodnotu distribuční funkce $F(x)$.



Funkce pro výpočet kvantilů tohoto rozdělení není v Excelu k dispozici. Pro výpočet $100P\%$ kvantilu Weibullova rozdělení můžeme použít vztah

$$x_P = \beta\left[-\ln(1-P)\right]^{1/\alpha}, \quad 0 < P < 1.$$

Studentovo rozdělení (t-rozdělení)

Náhodná veličina X má Studentovo rozdělení s parametrem n (počet stupňů volnosti), jestliže její hustota pravděpodobnosti má tvar

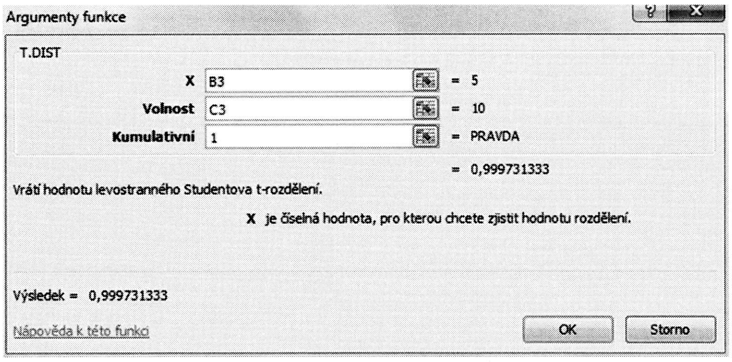
$$f(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{\pi n}}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \qquad -\infty < x < +\infty, \quad n \in N.$$

V Excelu se pro výpočet hodnot distribuční funkce používá funkce T.DIST. Argumenty funkce T.DIST mají následující význam:

X - x. Hodnota, ve které počítáme výraz $F(x)$ či $f(x)$.

Volnost - n. Parametr rozdělení, počet stupňů volnosti.

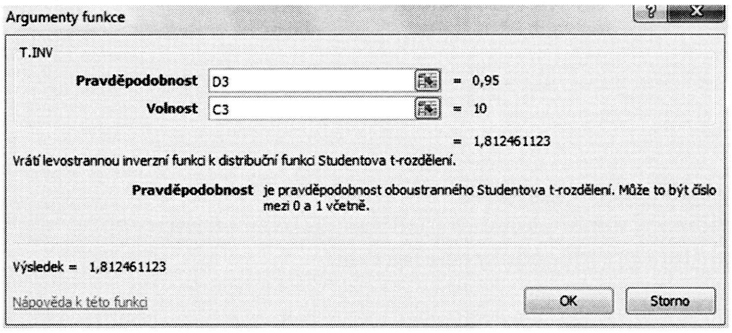
Kumulativní - NEPRAVDA pro hodnotu hustoty $f(x)$, PRAVDA pro hodnotu distribuční funkce $F(x)$.



Funkce pro výpočet kvantilů Studentova rozdělení má v Excelu název T.INV.

Prst - pravděpodobnost P pro hodnotu kvantilu x_P.

Volnost - n. Parametr rozdělení, počet stupňů volnosti.

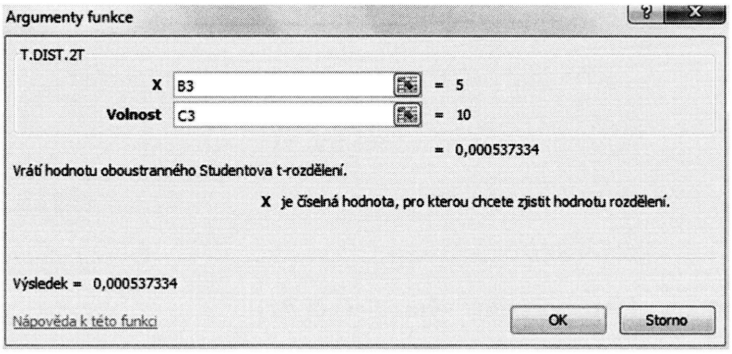


Další funkce
Excel disponuje dalšími funkcemi pro t-rozdělení.

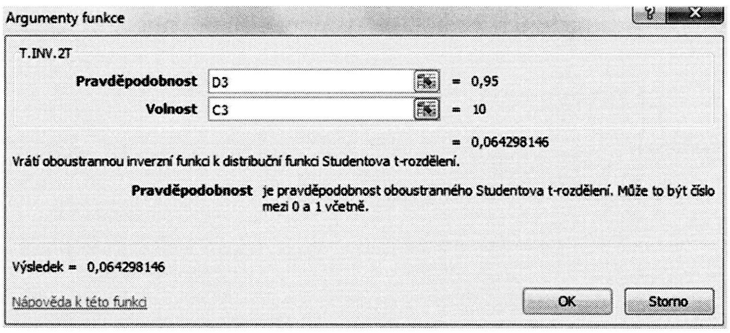
Jedná se o funkci T.DIST.2T. Tato funkce počítá pravděpodobnost $P(|X| > x)$, jde tedy o funkci kritických hodnot. Argumenty funkce T.DIST.2T mají následující význam:

X - x. Hodnota, ve které počítáme výraz $F(x)$ či $f(x)$.

Volnost - n. Parametr rozdělení, počet stupňů volnosti.



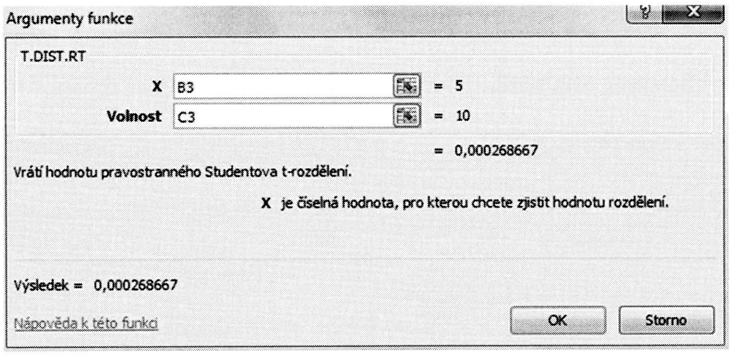
Další funkcí je T.INV.2T. Tato funkce počítá hodnotu kvantilu $x_{1-P/2}$. Argumenty funkce T.DIST.2T mají následující význam:



Pravděpodobnost - pravděpodobnost P pro hodnotu kvantilu $x_{1-P/2}$.

Volnost - n. Parametr rozdělení, počet stupňů volnosti.

Funkce T.DIST.RT počítá hodnotu $1-F(x)$, tedy doplněk distribuční funkce do jedné. Argumenty funkce T.DIST.RT mají následující význam:



X - x. Hodnota, ve které počítáme výraz $F(x)$ či $f(x)$.

Volnost - n. Parametr rozdělení, počet stupňů volnosti.

<u>Fischerovo-Schnedecorovo rozdělení (F rozdělení)</u>

Náhodná veličina X má Fischerovo-Schnedecorovo rozdělení s parametry m a n (počty stupňů volnosti), jestliže její hustota pravděpodobnosti má tvar

$$
\begin{aligned}
f(x) &= \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}\left(\frac{m}{n}\right)^{m/2} x^{m/2-1}\left(1+\frac{m}{n}x\right)^{-\frac{m+n}{2}}, \quad x>0,\ m\in N,\ n\in N,\\
&= 0, \qquad x \le 0.
\end{aligned}
$$

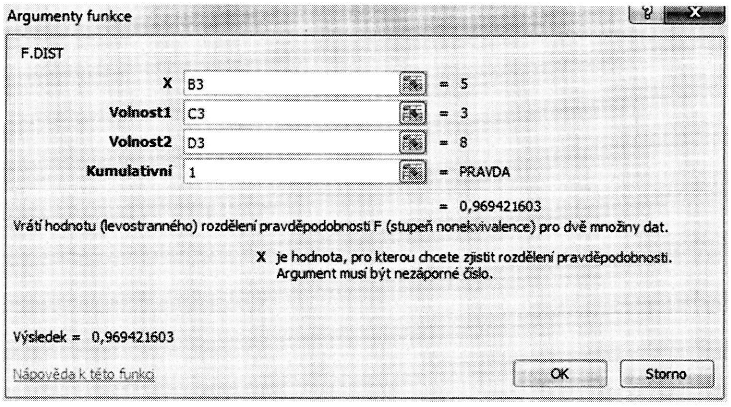
V Excelu se pro výpočet hodnot distribuční funkce používá funkce F.DIST. Argumenty funkce F.DIST mají následující význam:

X - x. Hodnota, ve které počítáme výraz $F(x)$ či $f(x)$.

Volnost1 - m. Parametr rozdělení, počet stupňů volnosti.

Volnost2 - n. Parametr rozdělení, počet stupňů volnosti.

Kumulativní - NEPRAVDA pro hodnotu hustoty $f(x)$, PRAVDA pro hodnotu distribuční funkce $F(x)$.

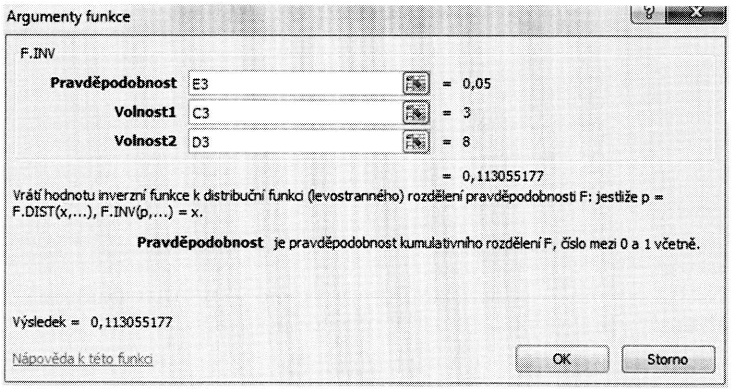


Pro výpočet hodnot kvantilů se používá funkce F.INV, která má následující parametry:

Pravděpodobnost - pravděpodobnost P pro hodnotu kvantilu x_p.

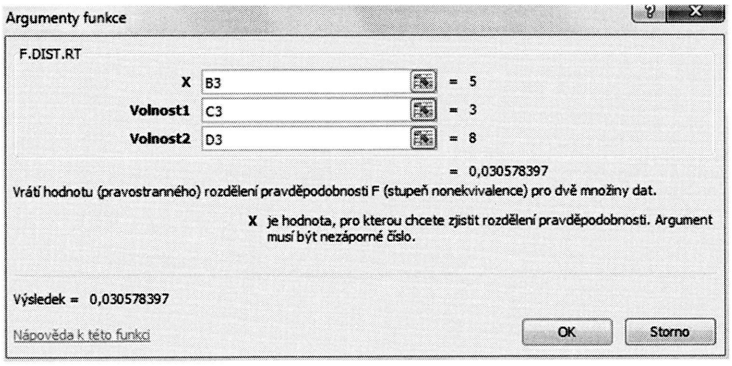
Volnost1 - m. Parametr rozdělení, počet stupňů volnosti.

Volnost2 - n. Parametr rozdělení, počet stupňů volnosti.



Dalšími funkcemi jsou F.DIST.RT a F.INV.RT. Funkce F.DIST.RT počítá hodnotu $1-F(x)$, tedy doplněk distribuční funkce do jedné, funkce F.INV.RT počítá hodnotu kvantilu x_{1-p}.

Argumenty obou funkcí mají následující význam:

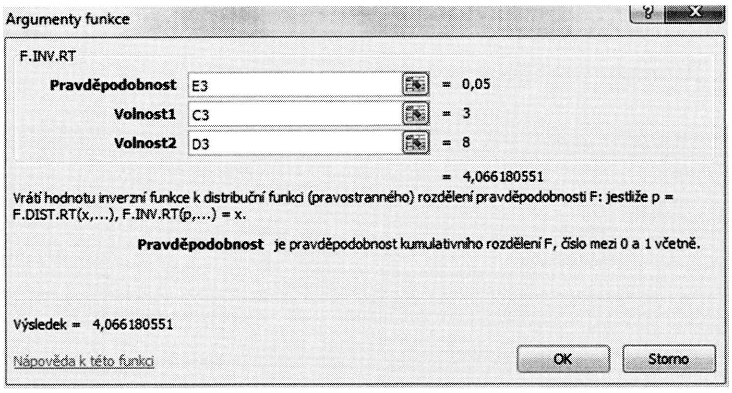


X - x. Hodnota, ve které počítáme výraz $F(x)$ či $f(x)$.

Volnost1 - m. Parametr rozdělení, počet stupňů volnosti.

Volnost2 - n. Parametr rozdělení, počet stupňů volnosti.

Inverzní funkcí k funkci předcházející je F.INV.RT s následujícími parametry.



Pravděpodobnost - pravděpodobnost P pro hodnotu kvantilu x_{1-P}.

Volnost1 - m. Parametr rozdělení, počet stupňů volnosti.

Volnost2 - n. Parametr rozdělení, počet stupňů volnosti.

Při výpočtu kvantilů Fischer-Schnedecorova rozdělení můžeme využít pro $0 < P < 1$ vztah

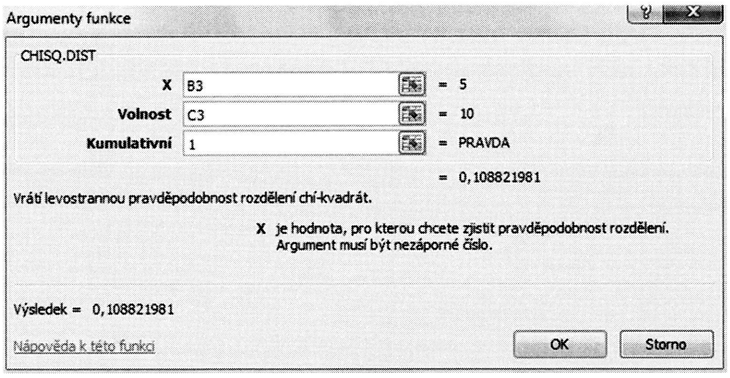
$$x_P(m,n) = \frac{1}{x_{1-P}(n,m)}.$$

Chí kvadrát rozdělení

Náhodná veličina *X* má chí-kvadrát rozdělení s parametrem *n* (počet stupňů volnosti), jestliže její hustota pravděpodobnosti má tvar

$$\begin{aligned} f(x) &= \frac{1}{2^{n/2}\Gamma\left(\frac{n}{2}\right)} x^{n/2-1} e^{-x/2}, \quad x > 0, \quad n \in N, \\ &= 0, \qquad\qquad\qquad\qquad\quad x \leq 0. \end{aligned}$$

V Excelu se pro výpočet hodnot distribuční funkce používá funkce CHISQ.DIST. Argumenty funkce CHISQ.DIST mají následující význam:

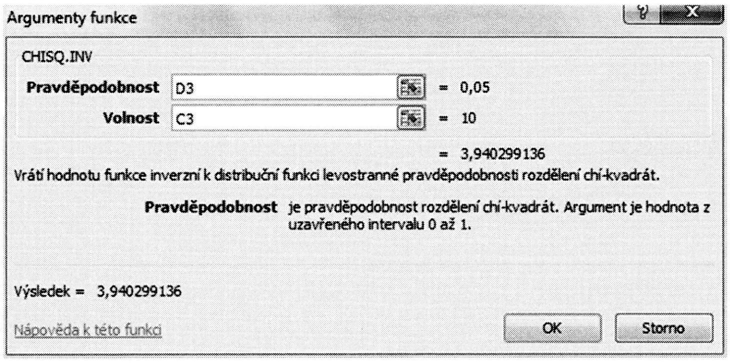


X - *x*. Hodnota, ve které počítáme výraz *F*(*x*) či *f*(*x*).

Volnost - *n*. Parametr rozdělení, počet stupňů volnosti.

Kumulativní - NEPRAVDA pro hodnotu hustoty *f*(*x*), PRAVDA pro hodnotu distribuční funkce *F*(*x*).

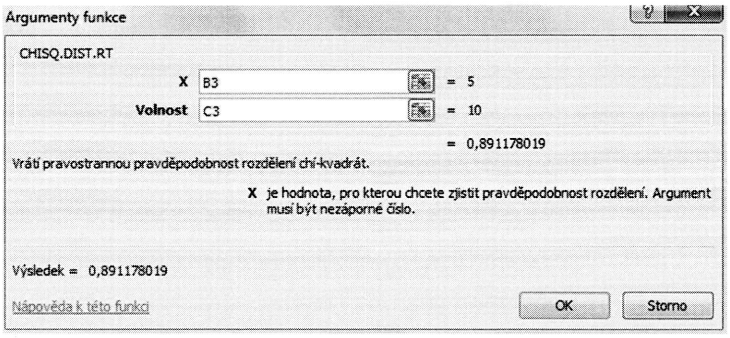
Pro výpočet hodnot kvantilů se používá funkce CHISQ.INV, která má následující parametry:



Pravděpodobnost - pravděpodobnost P pro hodnotu kvantilu x_P.

Volnost - n. Parametr rozdělení, počet stupňů volnosti.

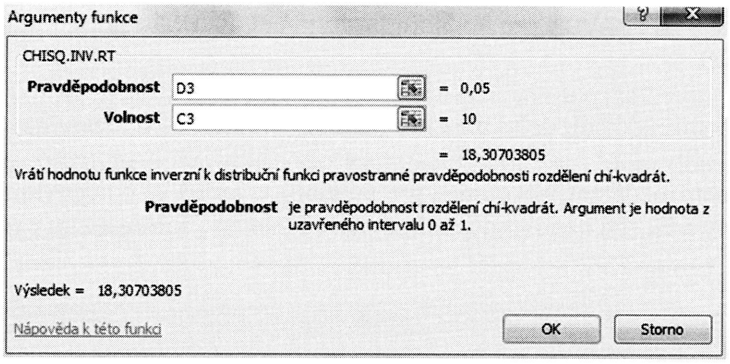
Dalšími funkcemi jsou CHISQ.DIST.RT a CHISQ.INV.RT. Funkce CHISQ.DIST.RT počítá hodnotu $1-F(x)$, tedy doplněk distribuční funkce do jedné. Její argumenty mají následující význam:



X - x. Hodnota, ve které počítáme výraz $F(x)$ či $f(x)$.

Volnost - n. Parametr rozdělení, počet stupňů volnosti.

Inverzní funkcí k funkci předchozí je CHISQ.INV.RT. Ta počítá hodnotu kvantilu x_{1-P}. Argumenty funkcí mají následující význam:



Pravděpodobnost - pravděpodobnost P pro hodnotu kvantilu x_{1-P}.

Volnost - n. Parametr rozdělení, počet stupňů volnosti.

Beta rozdělení (4 parametrické)

Náhodná veličina X má Beta rozdělení s parametry a, b, α, β, jestliže její hustota pravděpodobnosti má tvar

$$\begin{aligned} f(x) &= \frac{(x-a)^{\alpha-1}(b-x)^{\beta-1}}{B(\alpha,\beta)(b-a)^{\alpha+\beta-1}}, \quad a<x<b,\ \alpha>0,\ \beta>0,\ b>a, \\ &= 0, \qquad \text{jinak.} \end{aligned}$$

Pokud bychom položili $a=0$ a $b=1$, obdržíme „klasické“ dvouparametrické Beta rozdělení ve tvaru

$$\begin{aligned} f(x) &= \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}, \quad 0<x<1,\ \alpha>0,\ \beta>0, \\ &= 0, \qquad \text{jinak.} \end{aligned}$$

V Excelu se pro výpočet hodnot distribuční funkce používá funkce BETA.DIST. Argumenty funkce BETA.DIST mají následující význam:

X - x. Hodnota, ve které počítáme hodnotu distribuční funkce $F(x)$.

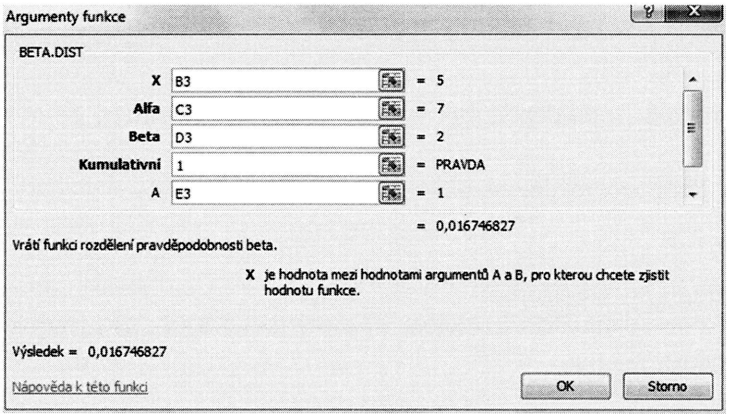
Alfa - α, parametr rozdělení.

Beta - β, parametr rozdělení.

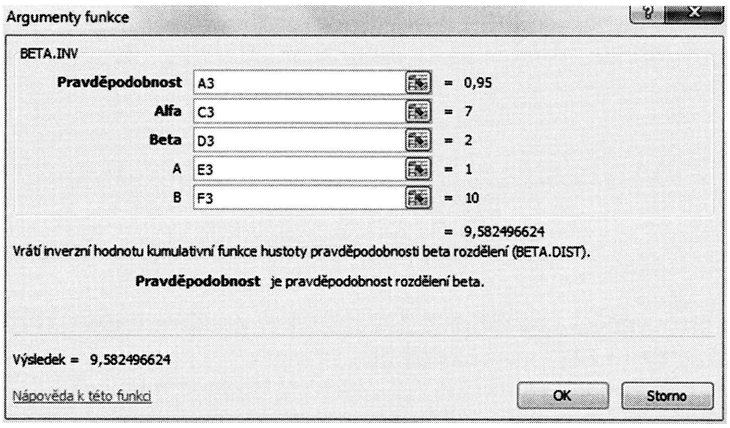
Kumulativní - NEPRAVDA pro hodnotu hustoty *f*(x), PRAVDA pro hodnotu distribuční funkce *F*(x).

A - a, parametr rozdělení, dolní mez pro hodnoty x. Jedná se o nepovinný argument.

B - b, parametr rozdělení, horní mez pro hodnoty x. Jedná se o nepovinný argument. Argumenty **A** a **B** jsou nepovinné, pokud nejsou zadány, automaticky platí **A** = 0 a **B** = 1.



Excel umožňuje i výpočet kvantilů beta rozdělení. Slouží k tomu funkce BETA.INV, jejíž parametry mají následující význam:



Pravděpodobnost - pravděpodobnost P pro hodnotu kvantilu x_P.

Alfa - α, parametr rozdělení.

Beta - β , parametr rozdělení.

A - a, parametr rozdělení, dolní mez pro hodnoty x. Jedná se o nepovinný argument.

B - b, parametr rozdělení, horní mez pro hodnoty x. Jedná se o nepovinný argument.

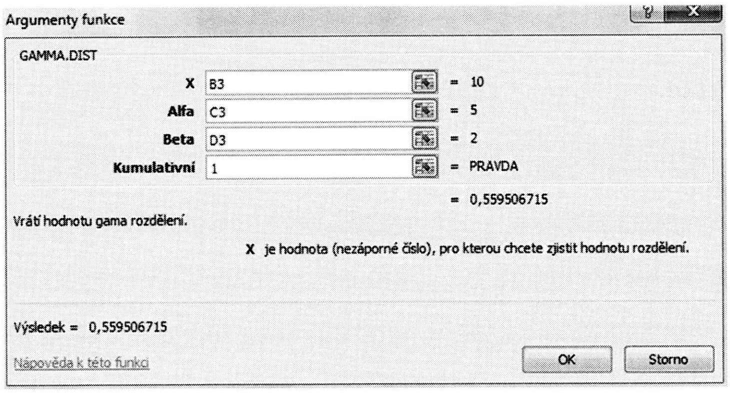
Pokud nejsou argumenty **A** a **B** zadány, automaticky platí **A** = 0 a **B** = 1.

Gama rozdělení

Náhodná veličina X má Gama rozdělení s parametry α a β , jestliže její hustota pravděpodobnosti má tvar

$$\begin{aligned} f(x) &= \frac{x^{\alpha-1}e^{-x/\beta}}{\beta^{\alpha}\,\Gamma(\alpha)}, \quad x>0,\ \alpha>0,\ \beta>0, \\ &= 0, \quad \text{jinak.} \end{aligned}$$

Pro výpočet hodnot distribuční funkce rozdělení gama se v Excelu používá funkce GAMMA.DIST. Argumenty funkce GAMMA.DIST mají následující význam:

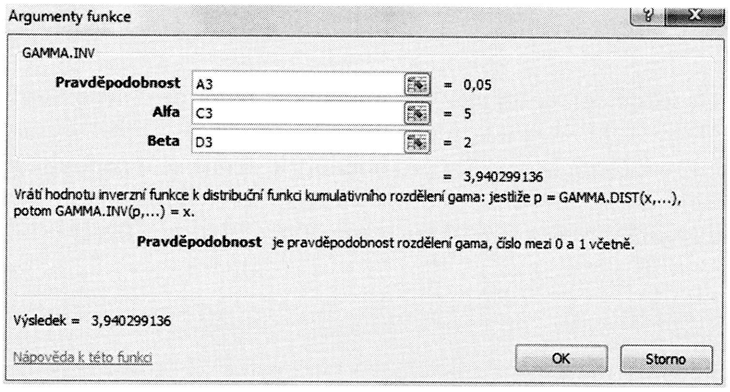


X - x. Hodnota, ve které počítáme hodnotu distribuční funkce $F(x)$.

Alfa - α , parametr rozdělení.

Beta - β , parametr rozdělení.

Kumulativní - NEPRAVDA pro hodnotu hustoty $f(x)$, PRAVDA pro hodnotu distribuční funkce $F(x)$.
Pro výpočet kvantilů se používá funkce GAMMA.INV.



Pravděpodobnost - pravděpodobnost P pro hodnotu kvantilu $\tilde{x}_P$.

Alfa - α, parametr rozdělení.

Beta - β, parametr rozdělení.

Pokud položíme parametr $\alpha = 1$, obdržíme exponenciální rozdělení ($\lambda = 1/\beta$).

Shrnutí syntaxe

Rozdělení	$F(x)$	$f(x)$	x_P
normální	=NORM.DIST(x;μ;σ;1)	=NORM.DIST(x;μ;σ;0)	=NORM.INV(P;μ;σ)
norm. normální	= NORM.S.DIST(x;1)	=NORM.S.DIST(x;0)	-
Log. normální	=LOGNORM.DIST(x;μ;σ;1)	=LOGNORM.DIST(x;μ;σ;0)	=LOGNORM.INV(P;μ;σ)
exponenciální	=EXPON.DIST(x;λ;1)	=EXPON.DIST(x;λ;0)	-
Weibullovo	=WEIBULL.DIST(x;α;β;1)	=WEIBULL.DIST(x;α;β;0)	-
t-rozdělení	=T.DIST(x;n;1)	=T.DIST(x;n;0)	=T.INV(P;n)
F-rozdělení	=F.DIST(x;n;m;1)	=F.DIST(x;n;m;0)	=F.INV(x;n;m)
Chí-kvadrát	=CHISQ.DIST(x;n;1)	=CHISQ.DIST(x;n;0)	=CHISQ.INV(P;n)
Beta	=BETA.DIST(x;α;β;1;a;b)	=BETA.DIST(x;α;β;0;a;b)	=BETA.INV(P;α;β;a;b)
Gama	=GAMMA.DIST(x;α;β;1)	=GAMMA.DIST(x;α;β;0)	=GAMMA.INV(P;α;β)

Další funkce

Rozdělení	$1-F(x)$	x_{1-P}	$P(\lvert X\rvert > x)$	$x_{1-P/2}$
t-rozdělení	=T.DIST.RT(x;n)	-	=T.DIST.2T(x;n)	=T.INV.2T(P;n)
F-rozdělení	=F.DIST.RT(x;m;n)	=F.INV.RT(P;m;n)	-	-
Chí-kvadrát	=CHISQ.DIST.RT(x;n)	=CHISQ.INV.RT(P;n)	-	-

KAPITOLA X

PŘÍLOHA

T A B U L K Y

u	$F(u)$	u	$F(u)$	u	$F(u)$
0,00	0,500 00	0,35	0,636 83	0,70	0,758 04
0,01	0,503 99	0,36	0,640 58	0,71	0,761 15
0,02	0,507 98	0,37	0,644 31	0,72	0,764 24
0,03	0,511 97	0,38	0,648 03	0,73	0,767 30
0,04	0,515 95	0,39	0,651 73	0,74	0,770 35

Ta

1.

2.

3.

4.

5.

6.

Tabulky

1. Distribuční funkce normovaného normálního rozdělení
2. Kvantily u_P normovaného normálního rozdělení
3. Kvantily χ_p^2 rozdělení χ^2 o ν stupních volnosti
4. Kvantily t_P t-rozdělení o ν stupních volnosti
5. Kvantily F_P F-rozdělení o ν_1 a ν_2 stupních volnosti
6. Kvantily rozdělení $D_{n;1-\alpha}$ pro Kolmogorovův-Smirnovův test pro jeden výběr

Tabulka 1 Distribuční funkce normovaného normálního rozdělení

u	$F(u)$	u	$F(u)$	u	$F(u)$	u	$F(u)$
0,00	0,500 00	0,35	0,636 83	0,70	0,758 04	1,05	0,853 14
0,01	0,503 99	0,36	0,640 58	0,71	0,761 15	1,06	0,855 43
0,02	0,507 98	0,37	0,644 31	0,72	0,764 24	1,07	0,857 69
0,03	0,511 97	0,38	0,648 03	0,73	0,767 30	1,08	0,859 93
0,04	0,515 95	0,39	0,651 73	0,74	0,770 35	1,09	0,862 14
0,05	0,519 44	0,40	0,655 42	0,75	0,773 77	1,10	0,864 33
0,06	0,523 92	0,41	0,659 10	0,76	0,776 37	1,11	0,866 50
0,07	0,527 90	0,42	0,662 76	0,77	0,779 35	1,12	0,868 64
0,08	0,531 88	0,43	0,666 40	0,78	0,782 30	1,13	0,870 76
0,09	0,535 86	0,44	0,670 03	0,79	0,785 24	1,14	0,872 86
0,10	0,539 83	0,45	0,673 64	0,80	0,788 14	1,15	0,874 93
0,11	0,543 80	0,46	0,677 24	0,81	0,791 03	1,16	0,876 98
0,12	0,547 76	0,47	0,680 80	0,82	0,793 89	1,17	0,879 00
0,13	0,551 72	0,48	0,684 39	0,83	0,796 73	1,18	0,881 00
0,14	0,555 67	0,49	0,687 93	0,84	0,799 55	1,19	0,882 98
0,15	0,559 62	0,50	0,691 46	0,85	0,802 34	1,20	0,884 93
0,16	0,563 56	0,51	0,694 97	0,86	0,805 11	1,21	0,886 86
0,17	0,567 49	0,52	0,698 47	0,87	0,807 85	1,22	0,888 77
0,18	0,571 42	0,53	0,701 94	0,88	0,810 57	1,23	0,890 65
0,19	0,575 35	0,54	0,705 40	0,89	0,813 27	1,24	0,892 51
0,20	0,579 26	0,55	0,708 84	0,90	0,815 94	1,25	0,894 35
0,21	0,583 17	0,56	0,712 26	0,91	0,818 59	1,26	0,896 17
0,22	0,587 06	0,57	0,715 66	0,92	0,821 21	1,27	0,897 96
0,23	0,590 95	0,58	0,719 04	0,93	0,823 81	1,28	0,899 73
0,24	0,594 83	0,59	0,722 40	0,94	0,826 39	1,29	0,901 47
0,25	0,598 71	0,60	0,725 75	0,95	0,828 94	1,30	0,903 20
0,26	0,602 57	0,61	0,729 07	0,96	0,831 47	1,31	0,904 90
0,27	0,606 42	0,62	0,732 37	0,97	0,833 98	1,32	0,906 58
0,28	0,610 26	0,63	0,735 65	0,98	0,836 46	1,33	0,908 24
0,29	0,614 09	0,64	0,738 91	0,99	0,838 91	1,34	0,909 88
0,30	0,617 91	0,65	0,742 15	1,00	0,841 34	1,35	0,911 49
0,31	0,621 72	0,66	0,745 37	1,01	0,843 75	1,36	0,913 09
0,32	0,625 52	0,67	0,748 57	1,02	0,846 14	1,37	0,914 66
0,33	0,629 30	0,68	0,751 75	1,03	0,848 50	1,38	0,916 21
0,34	0,633 07	0,69	0,754 90	1,04	0,850 83	1,39	0,917 74

Pro $u < 0$ jsou hodnoty distribuční funkce dány vztahem $F(-u) = 1 - F(u)$.

Tabulka 1 – pokračování

u	$F(u)$	u	$F(u)$	u	$F(u)$	u	$F(u)$
1,40	0,919 24	1,85	0,967 84	2,30	0,989 28	3,00	0,998 65
1,41	0,920 73	1,86	0,968 56	2,31	0,989 56	3,02	0,998 74
1,42	0,922 20	1,87	0,969 26	2,32	0,989 83	3,04	0,998 82
1,43	0,923 64	1,88	0,969 95	2,33	0,990 10	3,06	0,998 89
1,44	0,925 07	1,89	0,970 62	2,34	0,990 36	3,08	0,998 97
1,45	0,926 47	1,90	0,971 28	2,35	0,990 61	3,10	0,999 03
1,46	0,927 86	1,91	0,971 93	2,36	0,990 86	3,12	0,999 09
1,47	0,929 22	1,92	0,972 57	2,37	0,991 11	3,14	0,999 16
1,48	0,930 56	1,93	0,973 20	2,38	0,991 34	3,16	0,999 21
1,49	0,931 89	1,94	0,973 81	2,39	0,991 58	3,18	0,999 26
1,50	0,933 19	1,95	0,974 41	2,40	0,991 80	3,20	0,999 31
1,51	0,934 48	1,96	0,975 00	2,41	0,992 02	3,22	0,999 36
1,52	0,935 74	1,97	0,975 58	2,42	0,992 24	3,24	0,999 40
1,53	0,936 99	1,98	0,976 15	2,43	0,992 45	3,26	0,999 44
1,54	0,938 22	1,99	0,976 70	2,44	0,992 66	3,28	0,999 48
1,55	0,939 43	2,00	0,977 25	2,45	0,992 86	3,30	0,999 52
1,56	0,940 62	2,01	0,977 78	2,46	0,993 05	3,32	0,999 55
1,57	0,941 79	2,02	0,978 31	2,47	0,993 24	3,34	0,999 58
1,58	0,942 95	2,03	0,978 82	2,48	0,993 43	3,36	0,999 61
1,59	0,944 08	2,04	0,979 32	2,49	0,993 61	3,38	0,999 64
1,60	0,945 20	2,05	0,979 82	2,50	0,993 79	3,40	0,999 66
1,61	0,946 30	2,06	0,980 30	2,52	0,994 13	3,42	0,999 69
1,62	0,947 38	2,07	0,980 77	2,54	0,994 46	3,44	0,999 71
1,63	0,948 45	2,08	0,981 24	2,56	0,994 77	3,46	0,999 73
1,64	0,949 50	2,09	0,981 69	2,58	0,995 06	3,48	0,999 75
1,65	0,950 53	2,10	0,982 14	2,60	0,995 34	3,50	0,999 77
1,66	0,951 54	2,11	0,982 57	2,62	0,995 60	3,55	0,999 81
1,67	0,952 54	2,12	0,983 00	2,64	0,995 85	3,60	0,999 84
1,68	0,953 52	2,13	0,983 41	2,66	0,996 09	3,65	0,999 87
1,69	0,954 49	2,14	0,983 82	2,68	0,996 32	3,70	0,999 89
1,70	0,955 43	2,15	0,984 22	2,70	0,996 53	3,75	0,999 91
1,71	0,956 37	2,16	0,984 61	2,72	0,996 74	3,80	0,999 93
1,72	0,957 28	2,17	0,985 00	2,74	0,996 93	3,85	0,999 94
1,73	0,958 18	2,18	0,985 37	2,76	0,997 11	3,90	0,999 95
1,74	0,959 07	2,19	0,985 74	2,78	0,997 28	3,95	0,999 96

Pro $u < 0$ jsou hodnoty distribuční funkce dány vztahem $F(-u) = 1 - F(u)$.

Tabulka 1 – dokončení

u	$F(u)$	u	$F(u)$	u	$F(u)$	u	$F(u)$
1,75	0,959 94	2,20	0,986 10	2,80	0,997 44	4,00	0,999 97
1,76	0,960 80	2,21	0,986 45	2,82	0,997 60	4,05	0,999 97
1,77	0,961 64	2,22	0,986 79	2,84	0,997 74	4,10	0,999 98
1,78	0,962 46	2,23	0,987 13	2,86	0,997 88	4,15	0,999 98
1,79	0,963 27	2,24	0,987 45	2,88	0,998 01	4,20	0,999 99
1,80	0,964 07	2,25	0,987 78	2,90	0,998 13	4,25	0,999 99
1,81	0,964 85	2,26	0,988 09	2,92	0,998 25	4,30	0,999 99
1,82	0,965 62	2,27	0,988 40	2,94	0,998 36	4,35	0,999 99
1,83	0,966 38	2,28	0,988 70	2,96	0,998 46	4,40	0,999 99
1,84	0,967 12	2,29	0,988 99	2,98	0,998 56	4,45	1,000 00

Pro $u < 0$ jsou hodnoty distribuční funkce dány vztahem $F(-u) = 1 - F(u)$.

Tabulka 2 Kvantily normovaného normálního rozdělení u_P

P	u_P	P	u_P	P	u_P	P	u_P
0,50	0,000	0,75	0,674	0,950	1,645	0,975	1,960
0,51	0,025	0,76	0,706	0,951	1,655	0,976	1,977
0,52	0,050	0,77	0,739	0,952	1,665	0,977	1,995
0,53	0,075	0,78	0,772	0,953	1,675	0,978	2,014
0,54	0,100	0,79	0,806	0,954	1,685	0,979	2,034
0,55	0,126	0,80	0,842	0,955	1,695	0,980	2,054
0,56	0,151	0,81	0,878	0,956	1,706	0,981	2,075
0,57	0,176	0,82	0,915	0,957	1,717	0,982	2,097
0,58	0,202	0,83	0,954	0,958	1,728	0,983	2,120
0,59	0,228	0,84	0,994	0,959	1,739	0,984	2,144
0,60	0,253	0,85	1,036	0,960	1,751	0,985	2,170
0,61	0,279	0,86	1,080	0,961	1,762	0,986	2,197
0,62	0,305	0,87	1,126	0,962	1,774	0,987	2,226
0,63	0,332	0,88	1,175	0,963	1,787	0,988	2,257
0,64	0,358	0,89	1,227	0,964	1,799	0,989	2,290
0,65	0,385	0,900	1,282	0,965	1,812	0,990	2,326
0,66	0,412	0,905	1,311	0,966	1,825	0,991	2,366
0,67	0,440	0,910	1,341	0,967	1,838	0,992	2,409
0,68	0,468	0,915	1,372	0,968	1,852	0,993	2,457
0,69	0,496	0,920	1,405	0,969	1,866	0,994	2,512
0,70	0,524	0,925	1,440	0,970	1,881	0,995	2,576
0,71	0,553	0,930	1,476	0,971	1,896	0,996	2,652
0,72	0,583	0,935	1,514	0,972	1,911	0,997	2,748
0,73	0,613	0,940	1,555	0,973	1,927	0,998	2,878
0,74	0,643	0,945	1,598	0,974	1,943	0,999	3,090

Pro $P < 0{,}5$ jsou hodnoty kvantilů dány vztahem $u_P = -u_{1-P}$.

Tabulka 3 Kvantily χ_P^2 rozdělení χ^2 o ν stupních volnosti

ν	P						
	0,000 5	0,001	0,005	0,01	0,025	0,05	0,10
1	$0{,}0^{6}3\ 93$	$0{,}0^{5}1\ 57$	$0{,}0^{4}3\ 93$	$0{,}0^{3}1\ 57$	$0{,}0^{3}9\ 82$	$0{,}0^{2}3\ 93$	0,015 8
2	$0{,}0^{2}1\ 00$	$0{,}0^{2}2\ 00$	0,010 0	0,020 1	0,050 6	0,103	0,211
3	0,015 3	0,024 3	0,071 7	0,115	0,216	0,352	0,584
4	0,063 9	0,090 8	0,20 7	0,297	0,484	0,711	1,06
5	0,158	0,210	0,412	0,544	0,831	1,15	1,61
6	0,299	0,381	0,676	0,872	1,24	1,64	2,20
7	0,485	0,598	0,989	1,24	1,69	2,17	2,83
8	0,710	0,857	1,34	1,65	2,18	2,73	3,49
9	0,972	1,15	1,73	2,09	2,70	3,33	4,17
10	1,26	1,48	2,16	2,56	3,25	3,94	4,87
11	1,59	1,83	2,60	3,05	3,82	4,57	5,58
12	1,93	2,21	3,07	3,57	4,40	5,23	6,30
13	2,31	2,62	3,57	4,11	5,01	5,89	7,04
14	2,70	3,04	4,07	4,66	5,63	6,57	7,79
15	3,11	3,48	4,60	5,23	6,26	7,26	8,55
16	3,54	3,94	5,14	5,81	6,91	7,96	9,31
17	3,98	4,42	5,70	6,41	7,56	8,67	10,1
18	4,44	4,90	6,26	7,01	8,23	9,39	10,9
19	4,91	5,41	6,84	7,63	8,91	10,1	11,7
20	5,40	5,92	7,43	8,26	9,59	10,9	12,4
21	5,90	6,45	8,03	8,90	10,3	11,6	13,2
22	6,40	6,98	8,64	9,54	11,0	12,3	14,0
23	6,92	7,53	9,26	10,2	11,7	13,1	14,8
24	7,45	8,08	9,89	10,9	12,4	13,8	15,7
25	7,99	8,65	10,5	11,5	13,1	14,6	16,5
26	8,54	9,22	11,2	12,2	13,8	15,4	17,3
27	9,09	9,80	11,8	12,9	14,6	16,2	18,1
28	9,66	10,4	12,5	13,6	15,3	16,9	18,9
29	10,2	11,0	13,1	14,3	16,0	17,7	19,8
30	10,8	11,6	13,8	15,0	16,8	18,5	20,6

Tabulka 3 – dokončení

ν	P						
	0,90	0,95	0,975	0,99	0,995	0,999	0,999 5
1	2,71	3,84	5,02	6,63	7,88	10,8	12,1
2	4,61	5,99	7,38	9,21	10,6	13,8	15,2
3	6,25	7,81	9,35	11,3	12,8	16,3	17,7
4	7,78	9,49	11,1	13,3	14,9	18,5	20,0
5	9,24	11,1	12,8	15,1	16,7	20,5	22,1
6	10,6	12,6	14,4	16,8	18,5	22,5	24,1
7	12,0	14,1	16,0	18,5	20,3	24,3	26,0
8	13,4	15,5	17,5	20,1	22,0	26,1	27,9
9	14,7	16,9	19,0	21,7	23,6	27,9	29,7
10	16,0	18,3	20,5	23,2	25,2	29,6	31,4
11	17,3	19,7	21,9	24,7	26,8	31,3	33,1
12	18,5	21,0	23,3	26,2	28,3	32,9	34,8
13	19,8	22,4	24,7	27,7	29,8	34,5	36,5
14	21,1	23,7	26,1	29,1	31,3	36,1	38,1
15	22,3	25,0	27,5	30,6	32,8	37,7	39,7
16	23,5	26,3	28,8	32,0	34,3	39,3	41,3
17	24,8	27,6	30,2	33,4	35,7	40,8	42,9
18	26,0	28,9	31,5	34,8	37,2	42,3	44,4
19	27,2	30,1	32,9	36,2	38,6	43,8	46,0
20	28,4	31,4	34,2	37,6	40,0	45,3	47,5
21	29,6	32,7	35,5	38,9	41,4	46,8	49,0
22	30,8	33,9	36,8	40,3	42,8	48,3	50,5
23	32,0	35,2	38,1	41,6	44,2	49,7	52,0
24	33,2	36,4	39,4	43,0	45,6	51,2	53,6
25	34,4	37,7	40,6	44,3	46,9	52,6	54,9
26	35,6	38,9	41,9	45,6	48,3	54,1	56,4
27	36,7	40,1	43,2	47,0	49,6	55,5	57,9
28	37,9	41,3	44,5	48,3	51,0	56,9	59,3
29	39,1	42,6	45,7	49,6	52,3	58,3	60,7
30	40,3	43,8	47,0	50,9	53,7	59,7	62,2

Tabulka 4 Kvantily t_P rozdělení t o ν stupních volnosti

ν	P				
	0,90	0,95	0,975	0,99	0,995
1	3,078	6,314	12,706	31,821	63,657
2	1,886	2,920	4,303	6,965	9,925
3	1,638	2,353	3,182	4,541	5,841
4	1,533	2,132	2,776	3,747	4,604
5	1,476	2,015	2,571	3,365	4,032
6	1,440	1,943	2,447	3,143	3,707
7	1,415	1,895	2,365	2,998	3,499
8	1,397	1,860	2,306	2,896	3,355
9	1,383	1,833	2,262	2,821	3,250
10	1,372	1,812	2,228	2,764	3,169
11	1,363	1,796	2,201	2,718	3,106
12	1,356	1,782	2,179	2,681	3,055
13	1,350	1,771	2,160	2,650	3,012
14	1,345	1,761	2,145	2,624	2,977
15	1,341	1,753	2,131	2,602	2,947
16	1,337	1,746	2,210	2,583	2,921
17	1,333	1,740	2,110	2,567	2,898
18	1,330	1,734	2,101	2,552	2,878
19	1,328	1,729	2,093	2,539	2,861
20	1,325	1,725	2,086	2,528	2,845
21	1,323	1,721	2,080	2,518	2,831
22	1,321	1,717	2,074	2,508	2,819
23	1,319	1,714	2,069	2,500	2,807
24	1,318	1,711	2,064	2,492	2,797
25	1,316	1,708	2,060	2,485	2,787
26	1,315	1,706	2,056	2,479	2,779
27	1,314	1,703	2,052	2,473	2,771
28	1,313	1,701	2,048	2,467	2,763
29	1,311	1,699	2,045	2,462	2,756

Pro $P < 0{,}5$ jsou hodnoty kvantilů dány vztahem $t_P = -\,t_{1-P}$.

Tabulka 5a Kvantily $F_{0,95}$ rozdělení F o ν_1 a ν_2 stupních volnosti

ν_2	ν_1								
	1	2	3	4	5	6	7	8	9
1	161,450	199,500	215,710	224,580	230,160	233,990	236,770	238,880	240,540
2	18,513	19,000	19,164	19,247	19,296	19,330	19,253	19,371	19,385
3	10,128	9,552	9,277	9,117	9,014	8,941	8,887	8,845	8,812
4	7,709	6,944	6,591	6,388	6,256	6,163	6,094	6,041	5,999
5	6,608	5,786	5,410	5,192	5,050	4,950	4,876	4,818	4,773
6	5,987	5,143	4,757	4,534	4,387	4,284	4,207	4,147	4,099
7	5,591	4,737	4,437	4,120	3,972	3,866	3,787	3,726	3,677
8	5,318	4,459	4,066	3,838	3,688	3,581	3,501	3,438	3,388
9	5,117	4,257	3,863	3,633	3,482	3,374	3,293	3,230	3,179
10	4,965	4,103	3,708	3,478	3,326	3,217	3,136	3,072	3,020
11	4,844	3,982	3,587	3,357	3,204	3,095	3,012	2,948	2,796
12	4,747	3,885	3,490	3,259	3,106	2,996	2,913	2,849	2,896
13	4,667	3,806	3,411	3,179	3,025	2,915	2,832	2,767	2,714
14	4,600	3,739	3,344	3,112	2,958	2,848	2,764	2,699	2,646
15	4,543	3,682	3,287	3,056	2,901	2,791	2,707	2,641	2,588
16	4,494	3,634	3,239	3,007	2,852	2,741	2,657	2,591	2,538
17	4,451	3,592	3,197	2,965	2,810	2,699	2,614	2,548	2,494
18	4,414	3,555	3,160	2,928	2,773	2,661	2,577	2,510	2,456
19	4,381	3,522	3,127	2,895	2,740	2,628	2,544	2,477	2,423
20	4,351	3,493	3,098	2,866	2,711	2,599	2,514	2,447	2,393
21	4,325	3,467	3,073	2,840	2,685	2,573	2,488	2,421	2,366
22	4,301	3,443	3,049	2,817	2,661	2,549	2,464	2,397	2,342
23	4,279	3,422	3,028	2,796	2,640	2,528	2,442	2,375	2,320
24	4,260	3,403	3,009	2,776	2,621	2,508	2,423	2,355	2,300
25	4,242	3,385	2,991	2,759	2,603	2,490	2,405	2,337	2,282
26	4,225	3,369	2,975	2,743	2,587	2,475	2,388	2,321	2,266
27	4,210	3,354	2,960	2,728	2,572	2,459	2,373	2,305	2,250
28	4,196	3,340	2,947	2,714	2,558	2,445	2,359	2,291	2,236
29	4,183	3,328	2,934	2,701	2,545	2,432	2,346	2,278	2,223
30	4,171	3,316	2,922	2,690	2,534	2,421	2,334	2,266	2,211
40	4,085	3,232	2,839	2,606	2,450	2,336	2,249	2,180	2,124
60	4,001	3,150	2,758	2,525	2,368	2,254	2,167	2,097	2,040
120	3,920	3,072	2,680	2,447	2,290	2,175	2,087	2,016	1,959
∞	3,842	2,996	2,605	2,372	2,214	2,099	2,010	1,938	1,880

Tabulka 5a – dokončení

ν_2	ν_1									
	10	12	15	20	24	30	40	60	120	∞
1	241,88	243,91	245,95	248,01	249,05	250,09	251,14	252,20	253,25	254,32
2	19,396	19,413	19,429	19,446	19,454	19,462	19,471	19,479	19,487	19,496
3	8,786	8,745	8,703	8,660	8,639	8,617	8,594	8,572	8,549	8,527
4	5,964	5,912	5,858	5,803	5,774	5,746	5,717	5,688	5,658	5,628
5	4,735	4,678	4,619	4,558	4,527	4,496	4,464	4,431	4,398	4,365
6	4,060	4,000	3,938	3,874	3,842	3,808	3,774	3,740	3,705	3,669
7	3,637	3,575	3,511	3,445	3,411	3,376	3,340	3,304	3,267	3,230
8	3,347	3,284	3,218	3,150	3,115	3,079	3,043	3,005	2,967	2,928
9	3,137	3,073	3,006	2,937	2,901	2,864	2,826	2,787	2,748	2,707
10	2,978	2,913	2,845	2,774	2,737	2,700	2,661	2,621	2,580	2,538
11	2,854	2,788	2,719	2,646	2,609	2,571	2,531	2,490	2,448	2,405
12	2,753	2,687	2,617	2,544	2,506	2,466	2,426	2,384	2,341	2,296
13	2,671	2,604	2,533	2,459	2,420	2,380	2,339	2,297	2,252	2,206
14	2,602	2,534	2,463	2,388	2,349	2,308	2,266	2,223	2,178	2,131
15	2,544	2,475	2,404	2,328	2,288	2,247	2,204	2,160	2,114	2,066
16	2,494	2,425	2,352	2,276	2,235	2,194	2,151	2,106	2,059	2,010
17	2,450	2,381	2,308	2,230	2,190	2,148	2,104	2,058	2,011	1,960
18	2,412	2,342	2,269	2,191	2,150	2,107	2,063	2,017	1,968	1,917
19	2,378	2,308	2,234	2,156	2,114	2,071	2,026	1,980	1,930	1,878
20	2,348	2,278	2,203	2,124	2,083	2,039	1,994	1,946	1,896	1,843
21	2,321	2,250	2,176	2,096	2,054	2,010	1,965	1,917	1,866	1,812
22	2,297	2,226	2,151	2,071	2,028	1,984	1,938	1,890	1,838	1,783
23	2,275	2,204	2,128	2,048	2,005	1,961	1,914	1,865	1,813	1,757
24	2,255	2,183	2,108	2,027	1,984	1,939	1,892	1,842	1,790	1,733
25	2,237	2,165	2,089	2,008	1,964	1,919	1,872	1,822	1,768	1,711
26	2,220	2,148	2,072	1,990	1,946	1,901	1,853	1,803	1,749	1,691
27	2,204	2,132	2,056	1,974	1,930	1,884	1,836	1,785	1,731	1,672
28	2,190	2,118	2,041	1,959	1,915	1,869	1,820	1,769	1,714	1,654
29	2,177	2,105	2,028	1,945	1,901	1,854	1,806	1,754	1,698	1,638
30	2,165	2,092	2,015	1,932	1,887	1,841	1,792	1,740	1,684	1,622
40	2,077	2,004	1,925	1,839	1,793	1,744	1,693	1,637	1,577	1,509
60	1,993	1,917	1,863	1,748	1,700	1,649	1,594	1,534	1,467	1,389
120	1,911	1,834	1,751	1,659	1,608	1,554	1,495	1,429	1,352	1,254
∞	1,831	1,752	1,666	1,571	1,517	1,459	1,394	1,318	1,221	1,000

Tabulka 5b Kvantily $F_{0,975}$ rozdělení F o ν_1 a ν_2 stupních volnosti

ν_2	ν_1								
	1	2	3	4	5	6	7	8	9
1	647,790	799,500	864,160	899,580	921,850	937,110	948,220	956,660	963,280
2	38,506	39,000	39,165	39,248	39,298	39,331	39,355	39,373	39,387
3	17,443	16,044	15,439	15,101	14,885	14,735	14,624	14,540	14,473
4	12,218	10,649	9,979	9,605	9,365	9,197	9,074	8,980	8,905
5	10,007	8,434	7,764	7,388	7,146	6,978	6,853	6,757	6,681
6	8,813	7,260	6,599	6,227	5,988	5,820	5,696	5,600	5,523
7	8,073	6,542	5,890	5,523	5,285	5,119	4,995	4,899	4,823
8	7,571	6,060	5,416	5,053	4,817	4,652	4,529	4,433	4,357
9	7,209	5,715	5,078	4,718	4,484	4,320	4,197	4,102	4,026
10	6,937	5,456	4,826	4,468	4,236	4,072	3,950	3,855	3,779
11	6,724	5,256	4,630	4,275	4,044	3,881	3,759	3,664	3,588
12	6,554	5,096	4,474	4,121	3,891	3,728	3,607	3,512	2,436
13	6,414	4,965	4,347	3,996	3,767	3,604	3,483	3,388	3,312
14	6,298	4,857	4,242	3,892	3,663	3,501	3,380	3,285	3,209
15	6,200	4,765	4,153	3,804	3,576	3,415	3,293	3,199	3,123
16	6,115	4,687	4,077	3,729	3,502	3,341	3,219	3,125	3,049
17	6,024	4,619	4,011	3,665	3,438	3,277	3,156	3,061	2,985
18	5,978	4,560	3,954	3,608	3,382	3,221	3,100	3,005	2,929
19	5,922	4,508	3,903	3,559	3,333	3,172	3,051	2,956	2,880
20	5,872	4,461	3,859	3,515	3,289	3,128	3,007	2,913	2,837
21	5,827	4,420	3,819	3,475	3,250	3,090	2,969	2,875	2,798
22	5,786	4,383	3,783	3,440	3,215	3,055	2,934	2,839	2,763
23	5,750	4,349	3,751	3,408	3,184	3,023	2,902	2,808	2,731
24	5,717	4,319	3,721	3,379	3,155	2,995	2,874	2,779	2,703
25	5,686	4,291	3,694	3,353	3,129	2,969	2,848	2,753	2,677
26	5,659	4,266	3,670	3,329	3,105	2,945	2,824	2,729	2,653
27	5,633	4,242	3,647	3,307	3,083	2,923	2,802	2,707	2,631
28	5,610	4,221	3,626	3,286	3,063	2,903	2,782	2,687	2,611
29	5,588	4,201	3,607	3,267	3,044	2,884	2,763	2,669	2,592
30	5,568	4,182	3,589	3,250	3,027	2,867	2,746	2,651	2,575
40	5,424	4,051	3,463	3,126	2,904	2,744	2,624	2,529	2,452
60	5,286	3,925	3,343	3,008	2,786	2,627	2,507	2,412	2,334
120	5,152	3,805	3,227	2,894	2,674	2,515	2,395	2,299	2,222
∞	5,024	3,689	3,116	2,786	2,567	2,408	2,288	2,192	2,114

Tabulka 5b – dokončení

ν_2	ν_1									
	10	12	15	20	24	30	40	60	120	∞
1	968,63	976,71	984,87	993,10	997,25	1001,4	1005,6	1009,8	1014,0	1018,3
2	39,398	39,415	39,431	39,448	39,456	39,465	39,473	39,481	39,490	39,498
3	14,419	14,337	14,253	14,167	14,124	14,081	14,037	13,992	13,947	13,902
4	8,844	8,751	8,657	8,560	8,511	8,461	8,411	8,360	8,309	8,257
5	6,619	6,525	6,428	6,329	6,278	6,227	6,175	6,123	6,069	6,015
6	5,461	5,366	5,269	5,168	5,117	5,065	5,015	4,959	4,905	4,849
7	4,761	4,666	4,568	4,467	4,415	4,362	4,309	4,254	4,199	4,142
8	4,295	4,200	4,101	4,000	3,947	3,894	3,840	3,784	3,728	3,670
9	3,964	3,868	3,769	3,667	3,614	3,560	3,506	2,449	3,392	3,333
10	3,717	3,621	3,522	3,419	3,365	3,311	3,255	3,198	3,140	3,080
11	3,526	3,430	3,330	3,226	3,173	3,118	3,061	3,004	2,944	2,883
12	3,374	3,277	3,177	3,073	3,019	2,963	2,906	2,848	2,787	2,725
13	3,250	3,153	3,053	2,948	2,893	2,827	2,780	2,720	2,659	2,596
14	3,147	3,050	2,949	2,844	2,789	2,732	2,674	2,614	2,552	2,487
15	3,060	2,963	2,862	2,756	2,701	2,644	2,585	2,524	2,461	2,395
16	2,986	2,889	2,788	2,681	2,625	2,568	2,509	2,447	2,385	2,316
17	2,922	2,825	2,723	2,616	2,560	2,502	2,442	2,380	2,315	2,247
18	2,866	2,769	2,667	2,559	2,503	2,445	2,384	2,321	2,256	2,187
19	2,817	2,720	2,617	2,509	2,452	2,394	2,333	2,270	2,203	2,133
20	2,774	2,676	2,573	2,465	2,408	2,349	2,287	2,223	2,156	2,085
21	2,735	2,637	2,534	2,425	2,368	2,308	2,247	2,182	2,114	2,042
22	2,700	2,602	2,498	2,389	2,332	2,272	2,210	2,145	2,076	2,003
23	2,668	2,570	2,467	2,357	2,299	2,239	2,176	2,111	2,042	1,968
24	2,640	2,541	2,437	2,327	2,269	2,209	2,146	2,080	2,010	1,935
25	2,614	2,515	2,411	2,301	2,242	2,182	2,118	2,052	1,981	1,906
26	2,590	2,491	2,387	2,276	2,217	2,157	2,095	2,026	1,953	1,878
27	2,568	2,469	2,364	2,253	2,195	2,133	2,069	2,002	1,930	1,853
28	2,547	3,448	2,344	2,232	2,174	2,112	2,048	1,980	1,907	1,829
29	2,529	2,430	2,325	2,213	2,154	2,092	2,028	1,959	1,886	1,807
30	2,511	2,412	2,307	2,195	2,136	2,074	2,009	1,940	1,866	1,787
40	2,388	2,288	2,182	2,068	2,007	1,945	1,875	1,803	1,724	1,637
60	2,270	2,169	2,061	1,945	1,882	1,815	1,744	1,667	1,581	1,482
120	2,157	2,055	1,945	1,825	1,760	1,690	1,614	1,590	1,433	1,510
∞	2,048	1,945	1,833	1,709	1,640	1,566	1,484	1,588	1,268	1,000

∞
1018,3
39,498
13,902
8,257
6,015
4,849
4,142
3,670
3,333
3,080
2,883
2,725
2,596
2,487
2,395
2,316
2,247
2,187
2,133
2,085
2,042
2,003
1,968
1,935
1,906
1,878
1,853
1,829
1,807
1,787
1,637
1,482
1,510
1,000

Tabulka 5c Kvantily $F_{0,99}$ rozdělení F o ν_1 a ν_2 stupních volnosti

ν_2	ν_1								
	1	2	3	4	5	6	7	8	9
1	4052,2	4999,5	5403,3	5624,6	5763,7	5859,0	5928,3	5981,6	6022,5
2	98,503	99,000	99,166	99,249	99,299	99,332	99,356	99,374	99,388
3	34,116	30,817	29,457	28,710	28,237	27,911	27,672	27,489	27,345
4	21,198	18,000	16,694	15,977	15,522	15,207	14,976	14,799	14,659
5	16,258	13,274	12,060	11,392	10,967	10,207	10,456	10,289	10,158
6	13,745	10,925	9,780	9,148	8,746	8,466	8,260	8,102	7,976
7	12,246	9,547	8,451	7,847	7,460	7,191	6,993	6,840	6,719
8	11,259	8,469	7,591	7,006	6,632	6,371	6,178	6,029	5,911
9	10,561	8,022	6,992	6,422	6,057	5,802	5,613	5,467	5,351
10	10,044	7,559	6,552	5,994	5,636	5,386	5,200	5,057	4,942
11	9,646	7,206	6,217	5,668	5,316	5,069	4,886	4,745	4,632
12	9,330	6,927	5,953	5,412	5,064	4,821	4,640	4,499	4,388
13	9,074	6,701	5,739	5,205	4,862	4,620	4,441	4,302	4,191
14	8,862	6,515	5,564	5,035	4,695	4,456	4,278	4,140	4,030
15	8,683	6,359	5,417	4,893	4,556	4,318	4,132	4,005	3,895
16	8,531	6,226	5,292	4,773	4,437	4,202	4,026	3,890	3,780
17	8,400	6,112	5,185	4,669	4,336	4,102	3,927	3,791	3,682
18	8,285	6,013	5,092	4,579	4,248	4,015	3,841	3,705	3,597
19	8,185	5,926	5,010	4,500	4,171	3,939	3,765	3,631	3,523
20	8,096	5,849	4,938	4,431	4,103	3,871	3,699	3,564	3,457
21	8,017	5,780	4,874	4,369	4,042	3,812	3,640	3,506	3,398
22	7,945	5,719	4,817	4,313	3,988	3,758	3,578	3,453	3,346
23	8,881	5,664	4,765	4,264	3,939	3,710	3,539	3,406	3,299
24	7,823	5,614	4,718	4,218	3,895	3,667	3,496	3,363	3,256
25	7,770	5,568	4,676	4,177	3,855	3,627	3,457	3,324	3,217
26	7,721	5,526	4,637	4,140	3,818	3,591	3,421	3,288	3,128
27	7,667	5,488	4,601	4,106	3,785	3,558	3,388	3,256	3,149
28	7,636	5,453	4,568	4,074	3,754	3,528	3,358	3,226	3,120
29	7,598	5,421	4,538	4,045	3,725	3,500	3,330	3,198	3,092
30	7,536	5,390	4,510	4,018	3,699	3,474	3,305	3,173	3,067
40	7,314	5,179	4,313	3,828	3,514	3,291	3,124	2,993	2,888
60	7,077	4,977	4,126	3,649	3,339	3,119	2,953	2,823	2,719
120	6,851	4,787	3,949	3,480	3,174	2,956	2,792	2,663	2,559
∞	6,635	4,605	3,782	3,319	3,017	2,802	2,639	2,511	2,407

Tabulka 5c – dokončení

ν_2	ν_1									
	10	12	15	20	24	30	40	60	120	∞
1	6055,8	6106,3	6157,3	6208,7	6234,7	6260,7	6286,8	6313,0	6339,4	6366,0
2	99,399	99,416	99,432	99,449	99,458	99,466	99,474	99,483	99,491	99,501
3	27,229	27,052	26,872	26,690	26,598	26,505	26,411	26,316	26,221	26,125
4	14,546	14,374	14,198	14,020	13,929	13,838	13,745	13,652	13,558	13,463
5	10,051	9,888	9,722	9,553	9,467	9,379	9,291	9,202	9,112	9,020
6	7,874	7,718	7,559	7,396	7,313	7,229	7,143	7,057	6,969	6,880
7	6,620	6,469	6,314	6,155	6,074	5,992	5,908	5,824	5,737	5,650
8	5,814	5,667	5,515	5,359	5,279	5,198	5,116	5,032	4,946	4,859
9	5,257	5,111	4,962	4,808	4,729	4,649	4,567	4,483	4,398	4,311
10	4,849	4,706	4,558	4,405	4,327	4,247	4,165	4,082	3,997	3,909
11	4,539	4,397	4,251	4,099	4,021	3,941	3,860	3,776	3,690	3,603
12	4,296	4,155	4,010	3,858	3,781	3,701	3,619	3,536	3,449	3,361
13	4,100	3,960	3,815	3,665	3,587	3,507	3,425	3,341	3,255	3,165
14	3,939	3,800	3,656	3,505	3,427	3,348	3,266	3,181	3,094	3,004
15	3,805	3,666	3,522	3,372	3,294	3,214	3,132	3,047	2,960	2,868
16	3,691	3,553	3,409	3,259	3,181	3,101	3,018	2,933	2,845	2,753
17	3,593	3,455	3,312	3,162	3,084	3,003	2,921	2,835	2,746	2,653
18	3,508	3,371	3,227	3,077	2,999	2,919	2,835	2,749	2,660	2,566
19	3,434	3,297	3,153	3,003	2,925	2,844	2,761	2,674	2,584	2,489
20	3,368	3,231	3,088	2,938	2,859	2,779	2,695	2,608	2,517	2,421
21	3,310	3,173	3,030	2,880	2,801	2,720	2,636	2,548	2,457	2,360
22	3,258	3,121	2,978	2,827	2,749	2,668	2,583	2,495	2,403	2,306
23	3,211	3,074	2,931	2,781	2,702	2,620	2,536	2,447	2,354	2,256
24	3,168	3,032	2,889	2,738	2,659	2,577	2,492	2,404	2,310	2,211
25	3,129	2,993	2,850	2,699	2,620	2,538	2,453	2,364	2,270	2,169
26	3,094	2,958	2,815	2,664	2,585	2,503	2,417	2,327	2,233	2,132
27	3,062	2,926	2,783	2,632	2,552	2,470	2,384	2,294	2,198	2,097
28	3,032	2,896	2,753	2,602	2,522	2,440	2,354	2,263	2,167	2,064
29	3,005	2,869	2,726	2,574	2,495	2,412	2,325	2,234	2,138	2,034
30	2,979	2,843	2,700	2,549	2,469	2,386	2,299	2,208	2,111	2,006
40	2,801	2,665	2,522	2,369	2,288	2,203	2,114	2,019	1,917	1,805
60	2,632	2,496	2,352	2,198	2,115	2,029	1,936	1,836	1,726	1,601
120	2,472	2,336	2,192	2,035	1,950	1,860	1,763	1,656	1,533	1,381
∞	2,321	2,185	2,039	1,878	1,791	1,696	1,592	1,473	1,325	1,000

Tabulka 5d Kvantily $F_{0,995}$ rozdělení F o ν_1 a ν_2 stupních volnosti

ν_2	ν_1								
	1	2	3	4	5	6	7	8	9
1	16 211	20 000	21 615	22 500	23 056	23 437	23 715	23 925	24 091
2	198,50	199,00	199,17	199,25	199,30	199,33	199,36	199,37	199,39
3	55,552	49,799	47,467	46,195	45,392	44,838	44,434	44,126	43,882
4	31,333	26,284	24,259	23,155	22,456	21,975	21,622	21,352	21,139
5	22,785	18,314	16,350	15,556	14,940	14,513	14,200	13,961	13,772
6	18,635	14,544	12,917	12,028	11,464	11,073	10,786	10,566	10,391
7	16,236	12,404	10,882	10,050	9,155	9,155	8,885	8,678	8,514
8	14,688	11,042	9,597	8,805	8,302	7,964	7,694	7,496	7,339
9	13,614	10,107	8,717	7,956	7,471	7,134	6,885	6,693	6,541
10	12,826	9,427	8,081	7,343	6,872	6,545	6,303	6,116	5,968
11	12,226	8,912	7,600	6,881	6,422	6,102	5,865	5,682	5,537
12	11,754	8,510	7,226	6,521	6,071	5,757	5,525	5,345	5,202
13	11,374	8,187	6,926	6,234	5,791	5,482	5,253	5,076	4,935
14	11,060	7,922	6,680	5,998	5,562	5,257	5,031	4,857	4,717
15	10,798	7,701	6,476	5,803	5,372	5,071	4,847	4,674	4,536
16	10,575	7,514	6,303	5,638	5,212	4,913	4,692	4,521	4,384
17	10,384	7,354	6,156	5,497	5,075	4,779	4,559	4,389	4,254
18	10,218	7,215	6,028	5,375	4,956	4,663	4,445	4,276	4,141
19	10,073	7,094	5,916	5,268	4,853	4,561	4,345	4,177	4,043
20	9,944	6,987	5,818	5,174	4,762	4,472	4,257	4,090	3,956
21	9,830	6,891	5,730	5,091	4,681	4,393	4,179	4,013	3,880
22	9,727	6,806	5,652	5,017	4,609	4,323	4,109	3,944	3,812
23	9,635	6,730	5,582	4,950	4,544	4,259	4,047	3,882	3,750
24	9,551	6,661	5,519	4,890	4,486	4,202	3,991	3,826	3,695
25	9,475	6,598	5,462	4,835	4,433	4,150	3,939	3,776	3,645
26	9,406	6,541	5,409	4,785	4,384	4,103	3,893	3,730	3,599
27	9,342	6,489	5,361	4,740	4,340	4,059	3,850	3,688	3,557
28	9,284	6,440	5,317	4,698	4,300	4,020	3,811	3,649	3,519
29	9,230	6,396	5,276	4,659	4,262	3,983	3,775	3,613	3,483
30	9,180	6,355	5,239	4,623	4,228	3,949	3,742	3,580	3,451
40	8,828	6,066	4,976	4,374	3,986	3,713	3,509	3,350	3,222
60	8,495	5,795	4,729	4,140	3,760	3,492	3,291	3,134	3,008
120	8,179	5,539	4,497	3,921	3,548	3,285	3,087	2,933	2,808
∞	7,879	5,298	4,279	3,715	2,250	3,091	2,897	2,744	2,621

Tabulka 5d – dokončení

ν_2	ν_1									
	10	12	15	20	24	30	40	60	120	∞
1	24 224	24 426	24 630	24 836	24 940	25 044	25 148	25 253	25 359	25 465
2	199,40	199,42	199,43	199,45	199,46	199,47	199,47	199,48	199,49	199,51
3	43,686	43,387	43,085	42,778	42,622	42,466	42,308	42,149	41,989	41,829
4	20,967	20,705	20,438	20,167	20,030	19,892	19,752	19,611	19,468	19,325
5	13,168	13,384	13,146	12,903	12,780	12,656	12,530	12,402	12,274	12,144
6	10,250	10,034	9,814	9,589	9,474	9,358	9,241	9,122	9,002	8,879
7	8,380	8,176	7,968	7,754	7,645	7,535	7,423	7,309	7,193	7,076
8	7,811	7,015	6,814	6,608	6,503	6,396	6,288	6,177	6,065	5,951
9	6,417	6,227	6,033	5,832	5,729	5,625	5,519	5,410	5,300	5,188
10	5,847	5,661	5,471	5,274	5,173	5,071	4,966	4,859	4,750	4,639
11	5,418	5,236	5,049	4,855	4,756	4,654	4,551	4,445	4,337	4,226
12	5,086	4,906	4,721	4,530	4,432	4,331	4,228	4,123	4,015	3,904
13	4,820	4,643	4,460	4,270	4,173	4,073	3,970	3,866	3,758	3,647
14	4,603	4,428	4,247	4,059	3,961	3,862	3,760	3,655	3,547	3,436
15	4,424	4,250	4,070	3,883	3,786	3,687	3,585	3,480	3,372	3,260
16	4,272	4,099	3,921	3,734	3,638	3,539	3,437	3,332	3,224	3,112
17	4,142	3,971	3,793	3,607	3,511	3,412	3,311	3,206	3,097	2,984
18	4,031	3,860	3,683	3,498	3,402	3,303	3,201	3,096	2,987	2,873
19	3,933	3,763	3,587	3,402	3,306	3,208	3,106	3,000	2,891	2,776
20	3,847	3,678	3,502	3,318	3,222	3,123	3,022	2,916	2,806	2,690
21	3,771	3,602	3,427	3,243	3,147	3,049	2,947	2,841	2,730	2,614
22	3,703	3,535	3,360	3,176	3,081	2,982	2,880	2,774	2,663	2,546
23	3,642	3,475	3,300	3,117	3,021	2,922	2,820	2,713	2,602	2,484
24	3,587	3,420	3,246	3,062	2,967	2,868	2,765	2,659	2,546	2,428
25	3,537	3,370	3,196	3,013	2,918	2,819	2,716	2,609	2,496	2,377
26	3,492	3,325	3,152	2,969	2,873	2,774	2,671	2,563	2,450	2,330
27	3,450	3,284	3,110	2,928	2,832	2,733	2,630	2,522	2,408	2,287
28	3,412	3,246	3,073	2,890	2,794	2,695	2,592	2,483	2,369	2,247
29	3,377	3,211	3,038	2,855	2,759	2,660	2,557	2,448	2,333	2,210
30	3,344	3,179	3,006	2,823	2,727	2,628	2,524	2,415	2,300	2,176
40	3,117	2,953	2,781	2,598	2,502	2,402	2,296	2,184	2,064	1,932
60	2,904	2,742	2,571	2,387	2,290	2,187	2,079	1,962	1,834	1,688
120	2,705	2,544	2,373	2,188	2,089	1,984	1,871	1,747	1,606	1,431
∞	2,519	2,358	2,187	2,000	1,898	1,789	1,669	1,533	1,364	1,000

Tabulka 6 Kvantily rozdělení $D_{n;1-\alpha}$ při platnosti H_0 pro Kolmogorovův - Smirnovův test pro 1 výběr

n	$D_{n\,;\,0,9}$	$D_{n\,;\,0,95}$	$D_{n\,;\,0,99}$
1	0,950	0,975	0,995
2	0,776	0,842	0,929
3	0,636	0,708	0,829
4	0,565	0,624	0,734
5	0,509	0,563	0,669
6	0,468	0,519	0,617
7	0,436	0,483	0,576
8	0,410	0,454	0,542
9	0,387	0,430	0,513
10	0,369	0,409	0,489
11	0,352	0,391	0,468
12	0,338	0,375	0,449
13	0,325	0,361	0,432
14	0,314	0,349	0,418
15	0,304	0,338	0,404
16	0,295	0,327	0,392
17	0,286	0,318	0,380
18	0,279	0,309	0,371
19	0,271	0,301	0,361
20	0,265	0,294	0,352
21	0,259	0,287	0,344
22	0,253	0,281	0,337
23	0,247	0,275	0,330
24	0,242	0,269	0,323
25	0,238	0,264	0,317
26	0,233	0,259	0,311
27	0,229	0,254	0,305
28	0,225	0,250	0,300
29	0,221	0,246	0,295
30	0,218	0,242	0,290
31	0,214	0,238	0,285
32	0,211	0,234	0,281
33	0,208	0,231	0,277
34	0,205	0,227	0,273
35	0,202	0,224	0,269
36	0,199	0,221	0,265
37	0,196	0,218	0,262
38	0,194	0,215	0,258
39	0,191	0,213	0,255
40	0,189	0,210	0,252
41	0,187	0,208	0,249
42	0,185	0,205	0,246
43	0,183	0,203	0,243
44	0,181	0,201	0,241
45	0,179	0,198	0,238
46	0,177	0,196	0,235
47	0,175	0,194	0,233
48	0,173	0,192	0,231
49	0,171	0,190	0,228
50	0,170	0,188	0,226

LITERATURA

Anděl J.: *Matematická statistika*. Praha, SNTL/ALFA 1978.

Cipra, T.: *Analýza časových řad s aplikacemi v ekonomii*. Praha, SNTL/Alfa 1986, ISBN 04-012-86.

Čermák, V. , Vrabec, M.: *Teorie výběrových šetření*, 3. díl. VŠE Praha 1999, ISBN 80-245-0003-5.

Hátle J., Kahounová J.: *Úvod do teorie pravděpodobnosti*. Praha, SNTL 1987.

Hátle J., Likeš J.: *Základy počtu pravděpodobnosti a matematické statistiky*. Praha, SNTL Alfa 1987.

Hebák P., Kahounová J.: *Cvičení z teorie pravděpodobnosti*. Praha, Státní pedagogické nakladatelství 1983.

Hebák P., Kahounová J.: *Počet pravděpodobnosti v příkladech*. Praha, SNTL 1978.

Hindls, R., Hronová, S., Seger, J.: *Statistika pro ekonomy*. 5. vyd. Praha : Professional Publishing, 2003.

Likeš J., Machek J.: *Počet pravděpodobnosti*. Praha, SNTL 1991.

Likeš J., Laga J.: *Základní statistické tabulky*. Praha, SNTL 1978.

Marek a kol.: *Statistika pro ekonomy – aplikace*. Praha, Professional Publishing 2005.

Renyi A.: *Teorie pravděpodobnosti*. Praha, Academia 1972.

URL Český statistický úřad: http://www.czso.cz/

URL Česká národní banka: http://www.cnb.cz/